

Metadaten und Merkmale zur Verwaltung von persönlichen Musiksammlungen

Belegarbeit an der
Technischen Universität Dresden
September 2009

Thomas Gängler

Betreuer: Dipl.-Medieninf. Annett Mitschick
Hochschullehrer: Prof. Dr.-Ing. Klaus Meißner
Fachgebiet Multimediatechnik
Institut für Software- und Multimediatechnik
Fakultät Informatik



Erklärung

Hiermit erkläre ich, Thomas Gängler, die vorliegende Belegarbeit zum Thema
Metadaten und Merkmale zur Verwaltung von persönlichen Musiksammlungen
selbständig und ausschließlich unter Verwendung der im Quellenverzeichnis aufgeführten
Literatur- und sonstigen Informationsquellen verfasst zu haben.

Dresden, dem 10.09.2009 Unterschrift

Inhaltsverzeichnis

1	Einführung	8
2	Anwendungsfälle im privaten Umfeld	14
2.1	Ordnen	14
2.2	Orientieren	16
2.3	Suchen	21
2.4	Empfehlen	24
2.5	Wiedergabelisten- und Mixgenerierung	26
2.6	Verändern	29
2.7	Gemeinsam Benutzen	31
2.8	Fazit	33
3	Musikmerkmale	34
3.1	Audiosignaleigenschaften	37
3.1.1	Einfache Audiosignaleigenschaften	37
3.1.1.1	Physikalische Beschreibungen	37
3.1.1.2	Hörsinnesempfindungen	40
3.1.2	Angereicherte Audiosignaleigenschaften	46
3.1.2.1	Perzeptuelle Eigenschaften	46
3.1.2.2	Raum-zeitliche Struktureigenschaften	51
3.2	Taxonomien	55
3.2.1	Expertentaxonomien	56
3.2.2	Folksonomien	57
3.2.3	Ein dritter Ansatz	58
4	Musikmetadaten	59
4.1	Einordnungen von Musikmetadaten	59
4.2	Musikmetadatenstandards und -ontologien	63
4.2.1	Formatgebundene Spezifikationen	64
4.2.2	Formatunabhängige Spezifikationen	65
4.2.2.1	MPEG-Standards	66
4.2.2.2	Das <i>Music Ontology</i> -Rahmenwerk	69
4.3	Musikmetadatendienste	74
4.3.1	Dienste zur Katalogisierung von Audiosignalträgern	74
4.3.2	Dienste zur Musikempfehlung und -entdeckung	77
4.4	Fazit	81
5	Merkmalsextraktion	84
5.1	Audiosignalextraktion und -anreicherung	85
5.1.1	Audiosignalextraktion - Grundlagen	86

5.1.2	Zeit-Frequenz-Transformation	87
5.1.3	Perzeptionsmodelle	88
5.1.4	Statistiken	89
5.1.4.1	Direkte Statistiken	89
5.1.4.2	Spektralstatistiken	90
5.1.4.3	Perzeptionsstatistiken	91
5.1.5	Histogramme	92
5.1.5.1	Histogramme zur Rhythmusrepräsentation	92
5.1.5.2	Histogramme zur Tonhöhenrepräsentation	94
5.1.5.3	Histogramme zur Klangfarbenrepräsentation	95
5.1.6	Audiosignalanreicherung	96
5.2	Metadatenanalyse und -anreicherung	97
5.2.1	Metadatenextraktion von formatgebundenen Spezifikationen	97
5.2.2	Metadatenanreicherung durch Informationssuche	99
5.2.3	Klassifikation, Zuordnung und Ähnlichkeiten von Musikmetadaten	101
5.2.3.1	Multivariate Analysemethoden	101
5.2.3.2	Algorithmen des Maschinellen Lernens	105
5.2.3.3	Ontologien zur Klassifikation und Ähnlichkeitsanalyse	106
5.2.4	Persönliche Musikwissenbasis	107
5.3	Fazit	108
6	Zusammenfassung und Ausblick	110

Abbildungsverzeichnis

1.1	Wissenschaftliche Disziplinen, welche bei dem Musikinformationsretrieval mit wirken (nach [Ser08])	10
1.2	MIR Im Kontext andere Forschungsbereiche der Informationsverwaltungssysteme [Inf09]	10
1.3	Musikmerkmale von MIR [Fin04]	11
2.1	Anzeige von Coverbildern mittels <i>Cover Flow</i> von <i>iTunes</i>	16
2.2	Darstellung von Künstlern und Bands auf einer Landkarte	17
2.3	Künstler mit deren Geburtsdatum auf einer Zeitleiste [Rai08a]	17
2.4	Künstler und Bands in Beziehung zueinander	18
2.5	Musikauswahl durch Schieberegler oder 3-D-Netz	19
2.6	Auswahl der Stimmung in einem Feld	19
2.7	Künstler, ihre Verbindungen und Veröffentlichungen auf einer Zeitleiste .	20
2.8	Räumliche Visualisierung einer Musiksammlung nach den Merkmalen <i>Gesang, zuletzt gespielt</i> und <i>Perkussivität</i>	20
2.9	Eigenschaftenauswahl bei der räumlichen Visualisierung	21
2.10	Auswahl der Merkmale zur Ähnlichkeitsanalyse und Anordnung der Empfehlungsergebnisse	25
2.11	Pfadselektion innerhalb einer räumlichen Visualisierung [Lil08]	28
2.12	Kreisselektion innerhalb einer räumlichen Visualisierung [Lil08]	29
2.13	<i>Listen Game</i> : Auswahl der besten und schlechtesten Beschreibung eines Musikstückes [TLBL07]	30
2.14	<i>Digitales Ausleihen</i> mit der <i>Copyright Ontology</i> modelliert	33
3.1	Kategorien von Musikmerkmalen mit Unterbereichen und Beispielen (fliederfarben)	34
3.2	Schema des peripheren auditorischen Systems [Roe00]	35
3.3	Für die Musikwahrnehmung relevante Hirnareale auf der linken Hirnhälfte (mit Begriffslegende) [dIMHR05]	36
3.4	Die wesentlichen auditorischen Verarbeitungsblöcke der ASA (nach [Mar08b])	36
3.5	Projektion eines Punktes in gleichförmiger Kreisbewegung [Roe00] . . .	38
3.6	Eine komplexe periodische Schwingung	40
3.7	Wanderwelle entlang der (ausgerollten) Basilarmembran [dIMHR05] . . .	41
3.8	Ungefähre Frequenzbereiche entlang der Basilarmembran der Cochlea [Gol05]	42
3.9	Die ausgerollte Cochlea mit dem Bezug zu den Tonheitsgrößen Mel und Bark und zur Frequenz	42
3.10	Relation der Konzepte <i>wahrnehmbarer Anfangsmoment des Einschwingvorganges</i> , <i>Einschwingvorgang</i> , <i>Anklingphase</i> , <i>Abklingphase</i>	43

3.11	Referenzkurven von gleichen Lautstärkepegeln (ungefähre Lautheitsumrechnung) [FZ07]	44
3.12	Durchschnittlicher menschlicher Hörbereich mit den Abschnitten für Musik und Sprache [FZ07]	45
3.13	Klangsignal mit (a) fester Amplitude und (b) Amplitudenschwankungen [Vas05]	45
3.14	Tonhöhenreichweite verschiedene Musikinstrumente und Stimmlagen in Bezug auf eingeschlossene Oktavräume und Frequenzumfang (Auszug von [mis09h])	48
3.15	C-Dur-Tonleiter in einem Notensystem	48
3.16	Schraubenlinienrepräsentation der Helligkeit und Tonqualität der 12 chromatischen Töne [HW30]	49
3.17	Noten- und Pausewerte in der gebräuchlichen Notation der westlichen Musiktheorie [Hem01]	49
3.18	Spektrogramm eines Tom-Tom-Schlages [Hem01]	50
3.19	<i>Tonika</i> (T), <i>Dominante</i> (D) und <i>Subdominante</i> (S) von <i>F-Dur</i> in <i>Konsonanz</i> und <i>Dissonanz</i> [Sau07]	52
3.20	Gegenüberstellung der verschiedenen zeitstrukturierenden Merkmale der Musik [Qpa09]	53
3.21	Ein Ausschnitt der Hierarchie aus der Taxonomie von <i>MP3.com</i> [SCBG08]	56
4.1	Eine Übersicht der <i>low-level</i> -Beschreibungen von <i>MPEG-7 Audio</i> [Mar04]	67
4.2	Beispielhafte Verwendung des Zeitleistenkonzeptes [RA07b]	70
4.3	Das Ereignismodell der <i>Event Ontology</i> [RA07a]	71
4.4	Der Musikentstehungsprozess mit Konzepten und Beziehungen der <i>Music Ontology</i> beschrieben [GR07]	73
5.1	Musiksignalanalyserahmenwerk von Jehan [Jeh05]	84
5.2	<i>Rhythmus-Muster</i> (oben) und <i>Rhythmus-Histogramm</i> (unten) auf Basis von 60 Modulationsfrequenzen [Lid06]	93
5.3	Eine <i>Chromagramm</i> von vier Klängen (Klick, Clave, Snare-Drum, Violine) [Jeh05]	94
5.4	Ein Beispiel eines <i>Notenfolgen-Histogrammes</i> [Par07]	95
5.5	Ein <i>VRT-Spektrogramm</i> (mit einer logarithmischen Frequenzskala) eines Musikstückauszuges, bei dem nur ein Piano gespielt wird [Par07]	96
5.6	Grundstruktur eines (dreischichtigen) künstlichen neuronalen Netzes [BPEW08]	104

Tabellenverzeichnis

3.1	Auf <i>Konsonanz</i> hindeutende Frequenzverhältnisse (siehe [Sau07])	52
4.1	Abdeckungsstatistik der Datenbank für populäre Musik von <i>AMG</i> [Mac09c]	77

1 Einführung

Die Musik spielt und spielte im Leben der Menschen schon immer eine sehr bedeutende Rolle. Ein Leben ohne sie ist heutzutage kaum vorstellbar – ob es nun die permanente und oft unbewusste (Hintergrund-) *Musikberieselung*, der wir uns fast an jedem Ort aussetzen oder aussetzen müssen, oder die bewusst empfundene und wahrgenommene Darbietung oder Wiedergabe eines Musikstückes ist. Das Phänomen Musik ist allgegenwärtig, gesellschafts-, klassen- und kulturübergreifend, und eigentlich auch nie wirklich vollständig ergründet worden, sofern dies überhaupt mit Mitteln der Naturwissenschaften möglich ist. Es geht tief in die Entwicklung der Menschheit zurück, was nicht zuletzt die ältesten, archäologisch identifizieren, ca. 50.000 Jahre alten Musikinstrumente belegen (siehe [Wim04]).

Doch was ist Musik eigentlich genau? Womit lässt sie sich charakterisieren und beschreiben? Warum mögen wir überhaupt ein bestimmtes Musikstück oder einen bestimmten Künstler? – Besteht sie nur aus periodischen Luftdruckänderungen bzw. Schwingungen, welche mit einer Dauer, einer Frequenz, einer Amplitude und einer Klangfarbe beschrieben und von einem Gegenstand, der geschlagen, gerieben, gestrichen, in den geblasen oder durch den Luft gepumpt bzw. von einer Stimme oder einem elektronischen Instrument erzeugt wird [Hol05]? – Nein, Musik ist viel mehr. Sie ist von Menschen mit einer Mimik, einer Gestik, einem Aussehen und Empfindungen geschaffen, und kann von einem Hörer an einem konkreten Ort in einer bestimmten Zeit mit genau diesen vier Eigenschaften auch gehört werden, wobei die damit verbundenen Erlebnisse, Gefühle, Erinnerungen, Träume und Gedanken nicht dieselben sein müssen und sich oft auch stark von einander unterscheiden.

Musik wird nicht nur zur Formung einer Gesellschaft oder Kultur mitverwendet, sondern auch bei der Bildung der eigenen Identität eingebunden (z.B. durch die Repräsentation der persönlichen Musiksammlung), mit der sich ein Mensch in der Gesellschaft positioniert und über die er sich in dieser auf unterschiedlichen Wegen der Kommunikation austauscht, z.B. durch Gewähren von Anerkennung oder Ablehnung in bestimmten Subkulturen. Sie ist deshalb auch oft Teil von gesellschaftlichen und religiösen Riten oder ein Element in einer multimedialen Umgebung (z.B. Filmmusik). Letztendlich kann sie auch heilen, und den Geist in Ruhe versetzen (siehe [Hol05]).

An den Versuch die Schönheit der Kunst, von der die Musik ein Teil ist, in eine allgemeingültige Formel zu pressen, sind schon viele Menschen gescheitert, weil diese einfach in eine relativistische Auffassung eines jeden einzelnen Menschen zu fassen ist, und immer ein Rest einer Unbestimmtheit übrig bleibt, d.h. die Subjektivität überwiegt im Allgemeinen die Objektivität und etwas *Rätselhaftes* macht gerade in der Musik den Reiz aus (vgl. [Hol05]). So kann ein Hörer Gefallen an einem Musikstück finden, ohne dass über dieses etwas ausgesagt wird, außer dass es dem Hörer gefällt oder nicht. Die Kunst ist immer noch eine Reflexion eines Künstler in einer Situation eines gesellschaftlichen Kontextes, und wird vom Betrachter, im Fall der Musik der Hörer, wieder in die Gesellschaft zurück reflektiert (z.B. indem er seine Meinung dazu abgibt).

Der Reiz der Wissenschaftler besteht nun genau darin, die Unmöglichkeit der nicht trivialen Aufklärung des Phänomens Musik trotzdem zu ergründen (vgl. [Dow03]), um die Ursachen dieser, d.h. ihre Beschreibungen und Merkmale, festzulegen, und wie die Musik unter gewissen Umständen auf den Menschen einwirkt und von ihm wahrgenommen wird.

Einordnung des Aufgabenthemas In dieser Arbeit soll es deshalb im Speziellen um die genaue Betrachtung der Merkmale und Beschreibungen von Musik im Kontext der Verwaltung persönlicher Musiksammlungen gehen, da diese viel über die Persönlichkeit eines Menschen aussagen kann und er zu ihr eine große emotionale Beziehung aufbauen kann (vgl. [CJJ04]). In der heutigen Zeit verlagert sich eine Musiksammlung (gerade im Alltag) mehr und mehr in den Bereich der digitalen Medien und die Sinnesempfindung der Haptik (das Greifbare), welche durch analoge, d.h. herkömmliche, Tonträger, wie z.B. Schallplatten oder CDs, erzeugt werden, rückt mehr und mehr in den Hintergrund und der pure Konsum, also das Hören, immer mehr in den Vordergrund.

Aufgrund dieser Entwicklung ist es wichtig sich mit den Zugangsmöglichkeiten zu digitalen Musiksammlungen zu befassen, um (neue) Methoden zu entwickeln, welche dem Benutzer einen einfachen und intuitiven Zugang zu seiner persönlichen digitalen Musiksammlung ermöglichen und er diese erleben und bearbeiten kann. Deshalb befasst sich diese Arbeit auch im Wesentlichen mit den persönlichen (oft nicht kleinen) digitalen Musiksammlungen, welche sich als eine Sammlung von Musikdokumenten einer Person definieren lässt und von dieser genutzt und organisiert wird (siehe [CJJ04]). In dieser Musiksammlung wird davon ausgegangen, dass die enthaltenen Musikdokumente auch wirklich der Person gehören, d.h. dass diese zumindest die Rechte zur Wiedergabe dieser Musikstücke besitzt. Nicht zuletzt lassen sich heutzutage analoge Musiksammlungen mit einfachen Mitteln digitalisieren und ermöglichen einem Anwender dadurch auch neue Zugänge, Organisationsformen und Einsichten in das eigene Musikarchiv, welche auf Basis moderner Musikmerkmalsextraktionstechniken eine Vielzahl an Informationen aus den Abtastungen der Musikstücke gewinnen und aufbereiten, und darauf aufbauend über weitere Informationsquellen angereichert werden können.

Wissenschaftsbereich des Aufgabenthemas Der in dieser Belegarbeit genutzte interdisziplinäre Wissenschaftsbereich, welcher sich im Wesentlichen mit der (teil-)automatisierten Rückgewinnung von Beschreibungen und Merkmalen von Musik auf Basis von Musikdokumenten (d.h. Audiosignal und Metadaten) beschäftigt, um darauf erhaltenes oder generiertes Wissen in einer für den Anwender jeweilig geeigneten Form zugänglich zu machen und darzustellen, heißt *Musikinformatonsretrieval (MIR)* und kann auch *Musikinformatonsrückgewinnung* genannt werden. Viele wissenschaftliche Disziplinen und Teilbereiche tragen ihren Anteil zur Erforschung des hochkomplexen Prozesses – der Verarbeitung und des Umgangs mit der Musik durch den Menschen – bei, um mit den gewonnenen Erkenntnissen Systeme zu entwickeln, welche Benutzern große, digitale Musiksammlungen in vielfältiger Weise zugänglich machen können (vgl. [MKC06]). In Abbildung 1.1 ist eine grobe Übersicht der wichtigsten, beteiligten Wissenschaften zu sehen, wobei die Bereiche Musikkognition, Künstliche Intelligenz, Signalverarbeitung und Informationsrückgewinnung den Kernbereich bilden, aber dabei auch intensiv von dem Wissen der anderen Disziplinen Gebrauch machen.

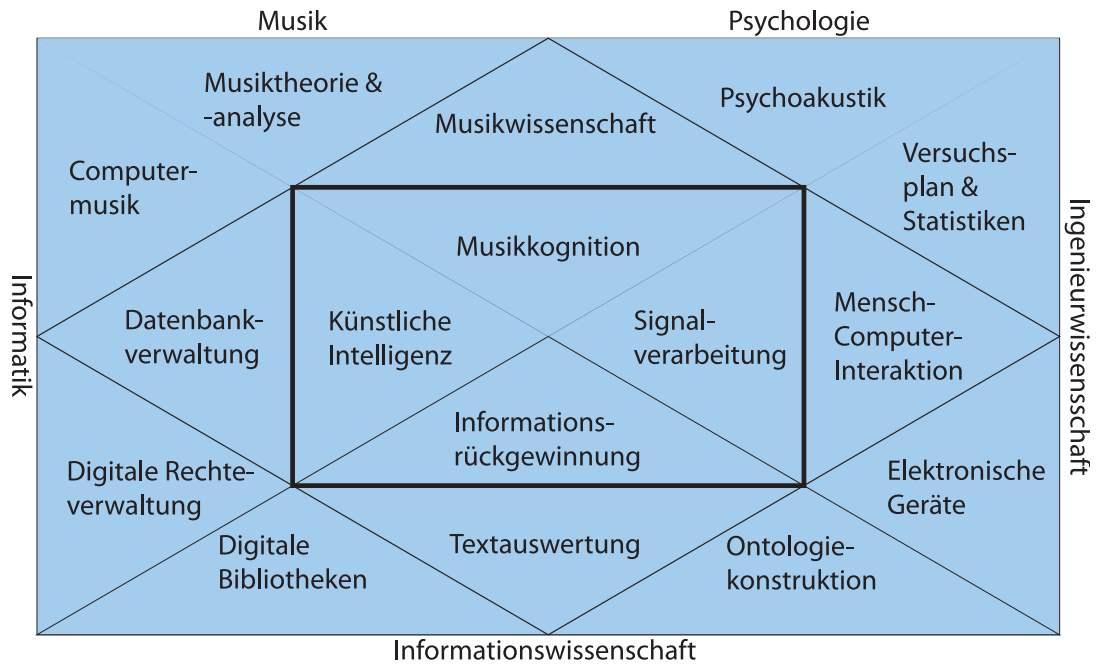


Abbildung 1.1: Wissenschaftliche Disziplinen, welche bei dem Musikinformationsretrieval mit wirken (nach [Ser08])

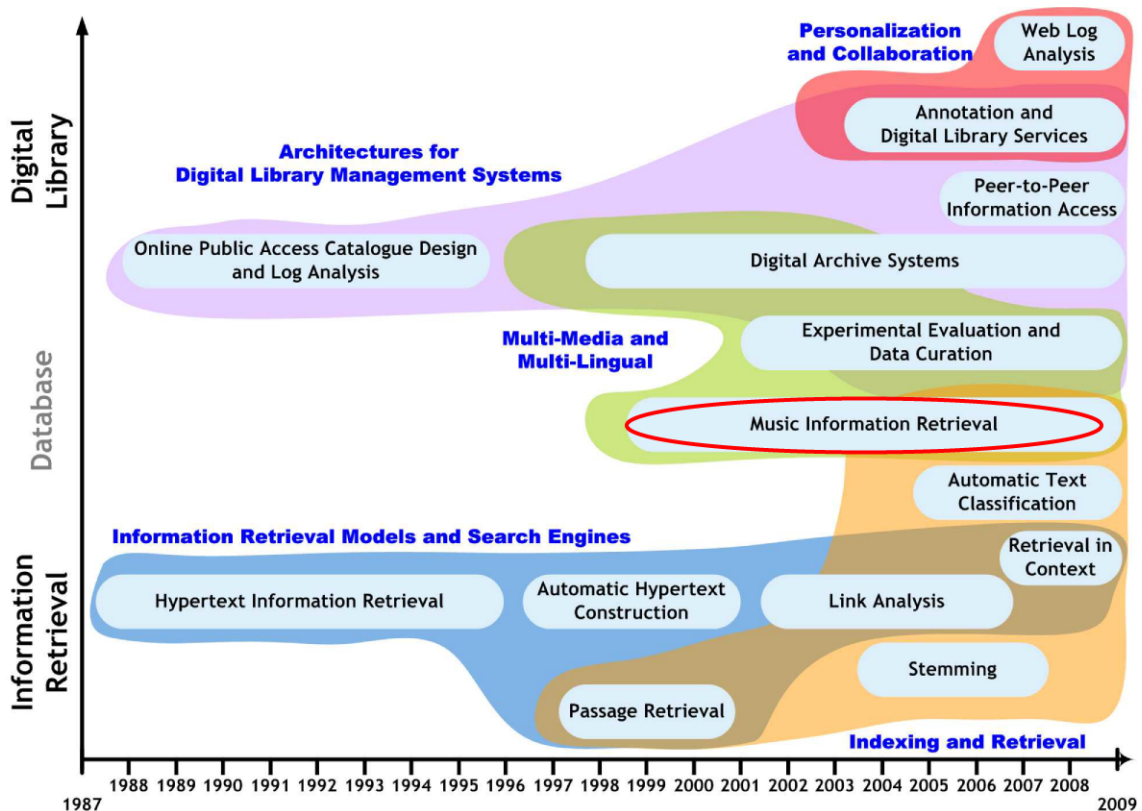


Abbildung 1.2: MIR Im Kontext andere Forschungsbereiche der Informationsverwaltungssysteme [Inf09]

Das *Musikinformatonsretrieval* ist gegen Ende der 1990er Jahre aufgekommen und bettet sich mitten in den Bereich der *Informationsverwaltungssysteme* (engl. *Information Management Systems, IMS*) ein. Dabei fließen auch die Forschungsergebnisse weiterer (teils älterer/ teils jüngerer) Retrievalkategorien, z.B. die Text- und Hypertextauswertung, und der Gestaltung und Architektur von digitalen Bibliotheken, z.B. die Kennzeichnung mit Schlagwörter (*Annotation*), mit in die Entwicklung von *MIR*-Systemen ein. In Abbildung 1.2 werden die einzelnen Forschungsbereiche und -gruppierungen in ihrer zeitlichen Entwicklung anschaulich dargestellt.

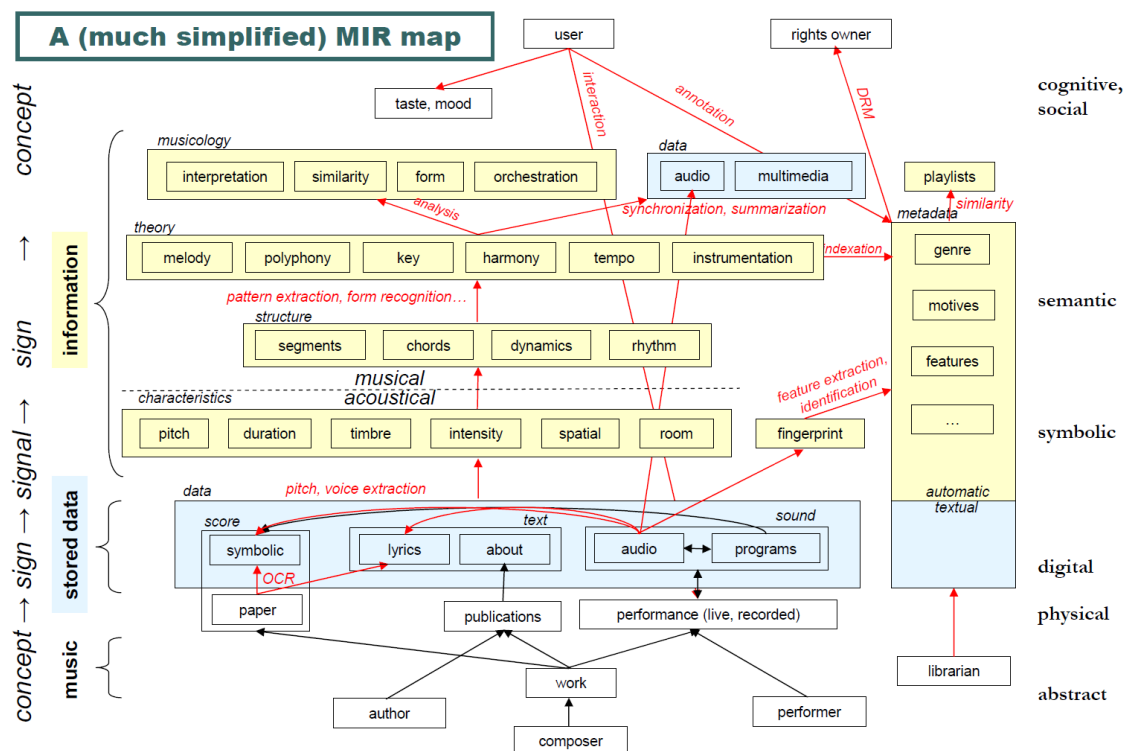


Abbildung 1.3: Musikmerkmale von MIR [Fin04]

Eine große Menge an Beschreibungen und Merkmalen der inhaltlichen und kontextuellen Erfassung von Musikinformationen kann bei dem *Musikinformatonsretrieval* entstehen. In Abbildung 1.3 ist eine allgemeine (unvollständige) Übersicht des Umfanges des Wissenraumes des *MIR* zu sehen. Der enorme Umfang drückt die Komplexität von Musik recht gut aus und zeigt auch, welches Wissen letztendlich nötig ist, um ungenaue verbale Anfragen von Anwendern zu beantworten bzw. die Informationen wieder bis zu diesem Niveau zu abstrahieren. Ein einfaches Beispiel ist die Auswahl von Musik nach einer bestimmten Stimmung, welche bei manueller Festlegung für einen Benutzer i.d.R. kein Problem ist, für eine (teil-)automatisierte Empfehlung einer Maschine aber schon.

***MIR* für persönliche digitale Musiksammlungen** Ein besonders aktueller Anwendungsbereich des *MIR* ist, wie bereits eingeführt, die Verwaltung persönlicher digitaler Musiksammlungen. Nicht nur durch populäre und an Bedeutung gewinnende Desкто-

panwendungen, wie z.B. *iTunes*¹, *Songbird*² oder *mufin player*³, erreichen die Forschungsergebnisse den Endanwender – nein, auch durch webbasierte Musikinformations- und -empfehlungsdienste, wie z.B. *Last.fm*⁴, *MusicBrainz*⁵ oder *All Music Guide*⁶, welche oftmals mit den Desk-topanwendungen zur Verwaltung von Musiksammlungen verknüpft werden und dadurch ein Informationsaustausch zwischen diesen Systemen stattfinden kann.

Die Musikdokumente, welche in diesem Bereich genutzt werden, sind im Wesentlichen audiosignalbasierte Dokumentenformate, wobei das Audiosignal in unkomprimierter (z.B. das *WAVE*-Dateiformat) oder komprimierter (z.B. das *MP3*-Dateiformat) Form kodiert wird. Die komprimierten Formate haben dabei die größere Verbreitung im Anwendungsbereich der persönlichen digitalen Musiksammlungen gefunden.

Aufbau der Arbeit Nachdem der Anwendungsbereich und das Forschungsgebiet dieser Belegarbeit vorgestellt worden sind, soll nun die grobe Unterteilung dieses Werkes erklärt werden, welche es sich zum Ziel gesetzt hat, die Musik- und musikrelevanten Metadaten und Merkmale zu untersuchen, die bei der Verwaltung von persönlichen Musiksammlungen zum Einsatz kommen können.

Zum Anfang werden die bedeutungsvollsten Anwendungsfälle, welche sich aus dem Themenbereich *persönliche digitale Musiksammlung* heraus entwickeln, in Kapitel 2 erläutert und durch aktuelle Lösungen dargestellt. Sie zeigen die umfangreichen Ansprüche auf, die an eine *Musikmerkmalsextraktion* und einem dazugehörigen *Musikmetadatenstandard* gestellt werden. Das allgemeine Ziel lautet dabei immer noch, einen Anwender, mit einer geeigneten Form und Menge an Musikmerkmalen und -metadaten, dabei zu unterstützen, neue Erfahrungen im Umgang mit der persönlichen Musiksammlung zu machen und nicht zuletzt die bekannten Verhaltensweisen mit einer analogen Musikkollektion in den digitalen Bereich umzusetzen.

Wenn ein Benutzer nun ein Musikstück seiner Musiksammlung anhört, startet ein hochkomplexer Prozess der Sinnesempfindung, Wahrnehmung und Verarbeitung (Kognition), in welchem eine Vielzahl von Assoziationen, Beschreibungen und Merkmale entstehen können, die er mit diesem Musikstück verbindet, d.h. wie er es bewertet und einordnet. Um diesen Ablauf zu verstehen, ist es wichtig, die grundlegenden Musikmerkmale genau darzustellen und ihre Zusammenhänge zu erläutern. So werden in Kapitel 3 neben den Audiosignaleigenschaften im physikalischen und psychoakustischen Bereich auch die daraus abgeleiteten Beschreibungen aus der Musiktheorie erklärt. Darüber hinaus folgt noch eine allgemeine Betrachtung von in der Musik relevanten Taxonomien, z.B. Musikgenretaxonomien.

Auf Basis der Musikmerkmale, aber auch weiterer kontextbezogener Informationen, wie z.B. Künstlername, Musikstücktitel oder Albumtitel, wurden und werden *Musikmetadatenformate* definiert und auf unterschiedlichste Weise eingesetzt. Einen Überblick über bedeutungsreiche und interessante Beschreibungen von *Musikmetadaten*, ihren Bezugsquellen und allgemeinen Gruppierungen, auf Basis von semantischen Gemeinsamkeiten,

¹<http://www.apple.com/de/itunes/>

²<http://getsongbird.com/>

³<http://www.mufin.com/de/software/mufinplayer>

⁴<http://www.lastfm.de/>

⁵<http://musicbrainz.org/>

⁶<http://www.allmusic.com/>

gibt Kapitel 4.

Mit den Erkenntnissen der ersten beiden Kapitel wird im darauf folgenden Kapitel 5 die Merkmalsextraktion für Musikinformationen auf Basis vieler, unterschiedlicher Informationsquellen beschrieben, welche sich in die zwei Teilbereiche der inhaltlichen und kontextuellen Wissenextraktion aufgliedert. Neben der allgemeinen Vorstellung grundlegender Techniken mit beispielhaften Realisierungen und Rahmenwerken, soll auch der Prozess beschrieben werden, der beim Auslesen der Musikedokumente beginnt und mit dem Abspeichern der gewonnenen und angereicherten Informationen endet. Dieser Ablauf ähnelt dabei im weitesten dem komplexen Wahrnehmungs- und Verarbeitungsprozess von Musiksignalen durch den Menschen.

In der Zusammenfassung (Kapitel 6) wird eine abschließende Betrachtung der gewonnenen Erkenntnisse dargestellt und aufgezeigt, dass nur eine auf den Benutzer optimierte Lösung zur vollständigen Zufriedenheit dessen führt. Der Ausblick soll dabei einmal mehr aufzeigen, dass, obwohl sich viele Anwender auch schon mit den jetzigen Programmen zur Verwaltung persönlicher Musiksammlungen mehr oder weniger zufrieden geben, sie eventuell umso mehr begeistert wären, wenn sie andere neue Zugänge nutzen können, welche ihren Wahrnehmungen besser entsprechen.

Abgrenzungen Diese Belegarbeit behandelt weder den grundlegenden Aufbau von Musikedokumenten, noch deren verschiedene Formate. Musikedokumente, die auf symbolischen Repräsentationen der Musik (z.B. Notenwerte) basieren, wie z.B. das *MIDI*-Dateiformat, werden in dieser Arbeit nicht untersucht. Im weiteren Verlauf dieser Ausarbeitung ist ein Musikedokument deshalb immer eine Datei eines Dateiformates, welches ein Audiosignal beinhaltet.

Durch die umfangreiche und tiefgründige Untersuchung der Anwendungsfälle im privaten Umfeld (siehe Kapitel 2) in Verbindung mit den Musikmerkmalen (siehe Kapitel 3), Musikmetadaten (siehe Kapitel 4) und der dazugehörigen Merkmalsextraktion (siehe Kapitel 5), ist es im Rahmen dieser Belegarbeit nicht möglich, ein aus den gewonnenen Erkenntnissen entworfenes Konzept zur Realisierung einer *Audioanalysekomponente* im Rahmen einer *Semantic Desktop*-Umgebung zu entwerfen und eine prototypische Implementierung dieses umzusetzen. In einer darauf aufbauenden Ausarbeitung, z.B. im Rahmen einer Diplomarbeit, kann das in diesem Aufsatz erarbeitete Wissen aber sehr gut angewendet werden, um eine *Musikmerkmalsextraktionskomponente*, deren Methoden, Ansätze und verwendbare Techniken in Kapitel 5 dieser Belegarbeit vorgestellt werden, zu entwickeln.

2 Anwendungsfälle im privaten Umfeld

Bei der Verwaltung von persönlichen Musiksammlungen können sich eine Vielzahl verschiedener Anwendungsfälle im privaten Umfeld ergeben. Die wichtigsten davon sollen in den folgenden Abschnitten dieses Kapitels vorgestellt werden, damit ein Überblick der vielfältigen Anforderungen an die Musikmerkmale und -metadaten entwickelt werden kann, welche in den darauffolgenden Kapiteln 3 und 4 ausführlich vorgestellt und analysiert werden.

Dabei ist es auch wichtig die Umgangsweisen mit einer physischen Musiksammlungen, mit denen einer digitalen Musiksammlung zu vergleichen bzw. diese darauf abzubilden und zu erweitern. Cunningham et al.[CJJ04] haben in ihrer Arbeit eine umfangreiche Untersuchung über die Organisationsformen für den Gebrauch von digitaler Musik, d.h. Musikedokumenten, durchgeführt und Anforderungen von physischen Musiksammlungen auf die digitalen Musiksammlungen angewandt bzw. neue eingeführt. Die in dieser Arbeit vorgestellten Anwendungsfälle im privaten Umfeld wurden von dieser und anderen Untersuchungen (z.B. [CBF06, LDJ07, LD04]) abgeleitet und gründlich auf die verwendeten oder benötigten Musikmerkmale und -metadaten erforscht. Sie heißen:

- *Ordnen* (siehe Abschnitt 2.1),
- *Orientieren* (siehe Abschnitt 2.2),
- *Suchen* (siehe Abschnitt 2.3),
- *Empfehlen* (siehe Abschnitt 2.4),
- *Wiedergabelisten- und Mixgenerierung* (siehe Abschnitt 2.5),
- *Verändern* (siehe Abschnitt 2.6),
- *Gemeinsam Benutzen* (siehe Abschnitt 2.7)

und können separat oder in einer Kombination bzw. Folge mehrerer Handlungen durch den Benutzer auf die digitale persönliche Musiksammlung angewandt werden.

Der Anwendungsfall *Wiedergeben* (von Musik) wurde dabei nicht separat betrachtet, da er nicht direkt zur Verwaltung zu zählen ist, aber als eine Folgehandlung von dem Benutzer ausgelöst werden kann und dabei auch verschiedene Gebrauchs- und Benutzungseigenschaften entstehen und gesammelt werden können (siehe Abschnitt 2.6).

2.1 Ordnen

Eine persönliche Musiksammlung kann nach vielen individuellen Kriterien geordnet werden. Generell hängt die grundsätzliche Strukturierung einer solchen Ordnung stark von

der Größe einer Musiksammlung ab.

Auf der einen Seite ist eine größere (physische) Musiksammlung (mehrere tausend Musikstücke) i.A. über mehrere Ort verteilt, was sich in digitalen Musiksammlungen zusätzlich in verschiedenen Sichten, d.h. Wiedergabelisten, widerspiegelt¹. Beispielhaft werden nun gebräuchliche Unterteilungen genannt², die sowohl auf einzelne Musikstücke aber auch auf Zusammenstellungen (wie z.B. Alben) angewandt werden und sich dabei in ihrer Größe unterscheiden³:

- **aktuelle Lieblingsstücke:** d.h. in letzter Zeit häufig gespielte Musikstücke; geringe Anzahl (ca. 10-20 Musikstücke bzw. 4-10 Zusammenstellungen)
- **allgemeine Favoriten:** d.h. über einen längeren Zeitraum häufig gespielte bzw. persönlich qualitativ wertvolle Objekte; etwas größere Anzahl (ca. 50-100 Musikstücke)
- **gelegentlich gespielte Einheiten:** i.A. der wesentliche Teil einer Musiksammlung, d.h. mit einer größeren (ca. 50-100 Musikzusammenstellungen) bis sehr großen Anzahl (je nach Größe der Musiksammlung), wenn sich dieser Bereich über die gesamte Musiksammlung erstreckt
- **gemeinsam genutzte Musiksammlungen:** z.B. der CD-Stapel im Wohnzimmer oder die Verteilung über ein lokales Netzwerk innerhalb einer Wohnung; variable Anzahl
- **Archiv:** enthält i.A. selten bis gar nicht gespielte Objekte (d.h. eine sehr große Anzahl), welche oft separat gelagert werden, oder es ist die gesamte Musiksammlung bzw. eine Sicherheitskopie davon

Auf der anderen Seite reicht bei einer größeren Sammlung die gebräuchliche Hierarchie mit Künstlernamen, Albumtitel, Musikstücktitel oft nicht mehr aus und es müssen weitere Ebenen eingeführt werden (z.B. Musikgenre oder Herkunftsland). Dabei unterscheiden sich die Ordnungsstrategien der einzelnen Unterteilungen bzw. Sichten stark voneinander. Wo bei den *allgemeinen Favoriten* das beliebteste Musikstück an vorderster Stelle ist⁴, ist es bei den aktuell häufig gespielten Einheiten, das mit der größten Wiedergabeanzahl. Bei den separaten Anordnungen mit einer größeren Objektanzahl können viele Faktoren mit in die Ordnungsstruktur einfließen (neben den schon erwähnten Eigenschaften der Hierarchiebildung), z.B. Kauf- oder Veröffentlichungsdatum (siehe [CJJ04]).

Zusätzlich zu den objektiven bzw. statistischen Einordnungen existieren auch sehr subjektive und individuelle Strategien, welche i.A. mehr eine spezielle Sicht der Musiksammlung verkörpern, als dass sie auf die gesamte Musiksammlung angewandt werden, d.h. dass das *Archiv* damit organisiert wird. Sie werden als *Musikkontext* (siehe Abschnitt 4.3.2 Paragraph „Rockanango Projekt“) bzw. *idiosynkratische Genre* (siehe [CJJ04]) bezeichnet und kategorisieren Musik nach ihrem Einsatzort und -zeit (z.B. „im Klub morgens um 5“ oder „abends im Wohnzimmer“) bzw. ihrer Verwendung (z.B. „Musik zum

¹Im Unterschied zur Ortsgebundenheit der Musikstücke in physischen Sammlungen

²Eine ausführliche Untersuchung dazu ist in [CJJ04] zu finden

³Im digitalen Bereich wird die Anzahl der enthaltenen Einheiten der jeweiligen Wiedergabelisten dabei häufig beschränkt

⁴Es muss dabei nicht einmal sehr häufig gespielt worden sein

Programmieren“ oder „Musik zum Auto fahren“). Hierbei werden viele inhaltliche Eigenschaften der Musik als Grundlage genutzt, z.B. Rhythmus, Melodie oder Grundtonart (siehe Abschnitt 3.1.2) und darüber hinaus aber auch Stimmungen, die durch diese Faktoren erzeugt werden, z.B. *energiegeladen*, *fröhlich* oder *entspannt*. Diese signifikanten Grundmuster (siehe Abschnitt 3.2) können mit Musikbeispielen hinterlegt werden und helfen dadurch der automatischen Einordnung neuer Musikstücke.

Bei der Benutzung von subjektiven Kategorisierungen in Archivhierarchien muss die Mehrfachbelegung von Eigenschaften (z.B. bei Musikgenres) beachtet werden⁵, d.h. eine eindeutige Eingliederung in die bestehende Sammlung ist dadurch ggf. nicht möglich. Mehrfachzuweisungen können dem Benutzer wiederum bessere und individualisierte Sichten ermöglichen. Unterstützt wird die Ordnung bzw. Einordnung i.A. durch eine Vielzahl softwaretechnischer Hilfsmittel, welche diese Prozesse (teil-)automatisieren (siehe Abschnitt 2.6).

Generell sollte jeder Benutzer die Individualisierung der Anordnungen in einem Profil, unter Beibehaltung der grundlegenden Hierarchie, hinterlegen können. Des Weiteren sollte die Möglichkeit bestehen, selbst erstellte *Musikkontexte* und Ordnungsstrategien auszutauschen, damit diese ggf. auf andere Inhalte angewandt werden können.

2.2 Orientieren

Mit Orientieren wird der Anwendungsfall abgedeckt, wie eine Musiksammlung durchsucht, in ihr navigiert und diese durchforscht werden kann. Wo bei den Such- und Empfehlungsanfragen (siehe Abschnitte 2.3 und 2.4) das Ergebnis von höchstem Interesse ist, ist es hier der Weg. So kann eine orientierende Suche, d.h. wenn man die Informationen zur eindeutigen Identifikation eines Musikstückes oder einer Zusammenstellung nicht genau kennt, auch vor dem eigentlichen Ziel abgebrochen werden, wenn der Suchende ein interessantes Objekt gefunden hat. Dabei kann er vom eigentlichen Suchpfad auch völlig abgekommen sein.



Abbildung 2.1: Anzeige von Coverbildern mittels *Cover Flow* von iTunes

Eine gewohnte Form des Orientierens ist das Durchsuchen eines CD- oder Schallplattenstapels, welches in Anwendungen zur Verwaltung von Musiksammlungen durch eine interaktive Anzeige der Coverbilder (engl. *Cover Flow*) realisiert ist (siehe Abbildung

⁵Sofern sie von der genutzten *Musikmetadatenpezifikation* zulässig ist, siehe Abschnitt 4.2

2.1). Mit dieser Darstellung kann ein Benutzer durch seine Musiksammlung oder Teile dieser effektiv blättern und sieht sofort die Frontansicht oder ggf. die Rückansicht des Mediums. Dies ist ein großer Vorteil gegenüber der physischen CD-Sammlung, bei der man sich i.A. zuerst den CD-Rücken (schmale Seite) anschaut, auf dem weniger und oft auch nur textuelle Informationen zu bekommen sind (vgl. [CJJ04]).

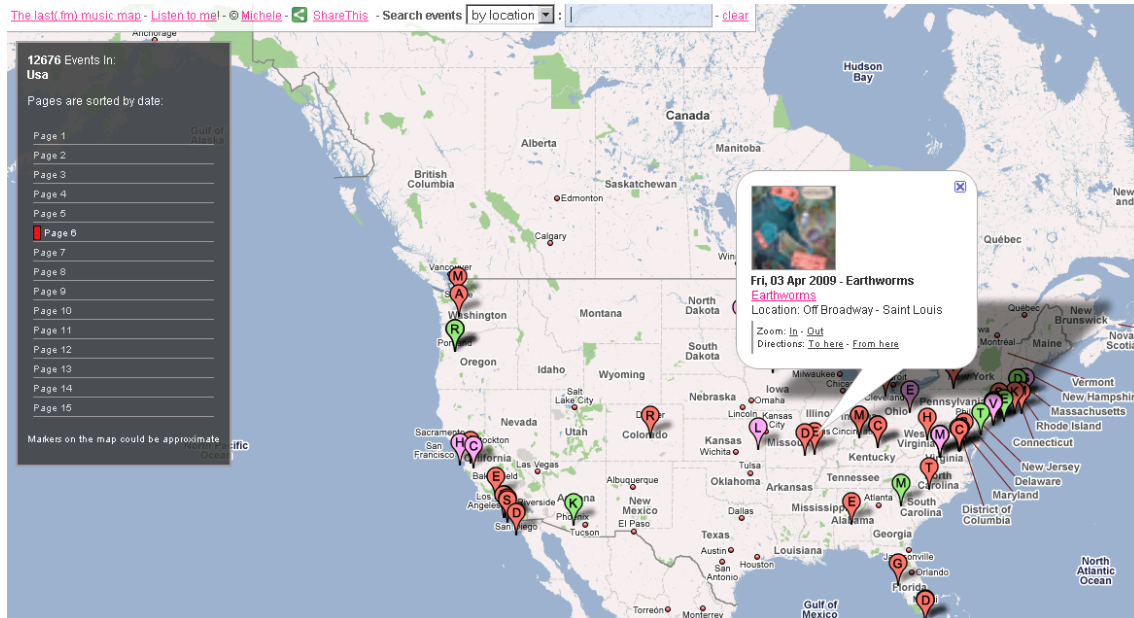


Abbildung 2.2: Darstellung von Künstlern und Bands auf einer Landkarte⁶

Generell können beim Navigieren alle möglichen *Metadatenfelder* zum Einsatz kommen (siehe Abschnitt 4.1 Paragraph „Metadatenfeldergruppierungen“). Die *Facettennavigation* (engl. *Faceted Browsing*[Rai08a]) ermöglicht es dabei dem Benutzer eine Einschränkung über verschiedene Kategorien und Merkmale vorzunehmen, wobei die einfachste Form die Auswahl von jeweiligen Eigenschaften aus Auflistungen, d.h. einfache Textlisten, der vorhandenen Kategorie- und Merkmalsentitäten ist. Intuitive Ansätze hingegen versuchen die verfügbaren Informationen visuell anspruchsvoller zu repräsentieren, wie das im Folgenden beschriebene Beispiel zeigen wird.

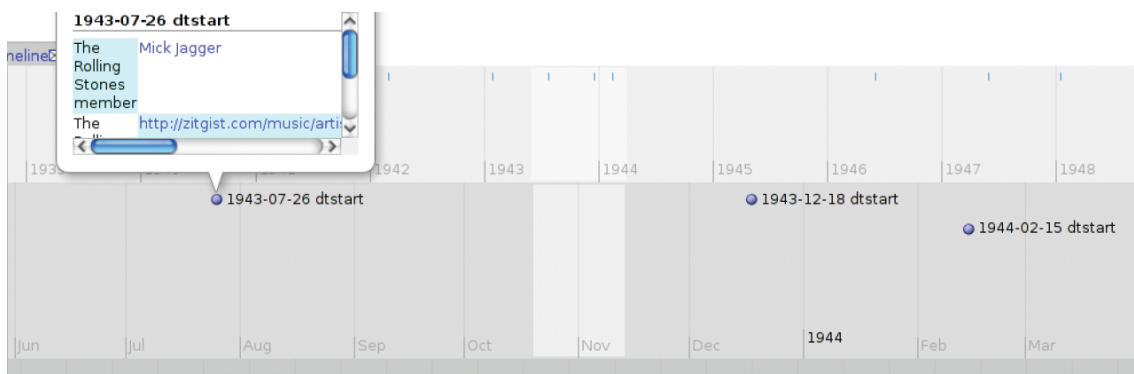


Abbildung 2.3: Künstler mit deren Geburtsdatum auf einer Zeitleiste [Rai08a]

⁶Ausschnitt aufgenommen von <http://www.lastmusicmap.com/#>

Ein Anwender kann sich z.B. seine Musiksammlung oder Ausschnitte davon auf einer Landkarte anzeigen lassen (siehe Abbildung 2.2), d.h. z.B. nach Herkunft der Künstler oder dem Ort der Veröffentlichung. Dabei wählt er ein paar Exemplare aus, indem er z.B. ein Gebiet in den USA markiert, und lässt sich diese wiederum auf einer Zeitleiste anzeigen (siehe Abbildung 2.3). Nun interessiert sich der Benutzer besonders für die musikalische Werke, die zwischen 1970 und 1975 veröffentlicht wurden. Letztendlich möchte er noch gern wissen, wie die Künstler dieser Werke zueinander in Beziehung stehen und wechselt dafür wieder in eine andere Sicht. Die Suche endet durch die Auswahl eines Künstler und allen mit ihm in Verbindung stehenden Künstlern, die mit ihm zusammen in dieser Zeit ein Musikstück aufgenommen haben (siehe beispielhaft Abbildung 2.4; wobei die Relationen auf Ähnlichkeitsbeziehungen beruhen). Bei diesem Beispiel wurden ausschließlich *redaktionelle* und kulturelle Eigenschaften verwendet (siehe Abschnitt 4.1 Paragraph „Metadatentypen“).

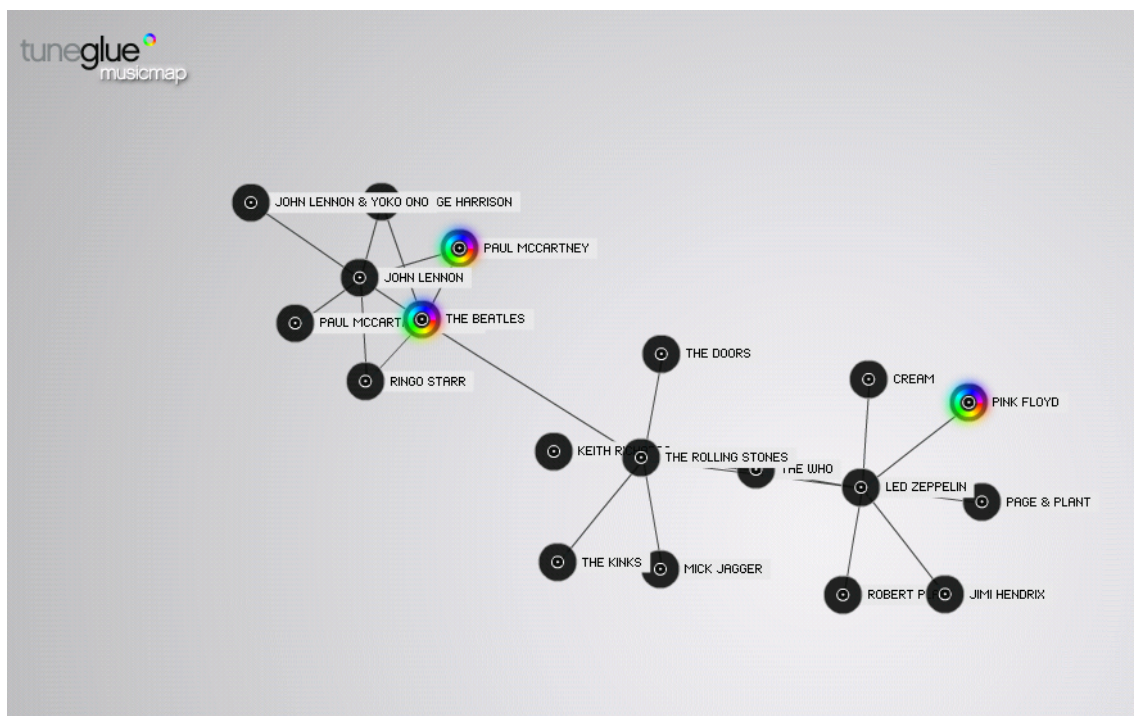
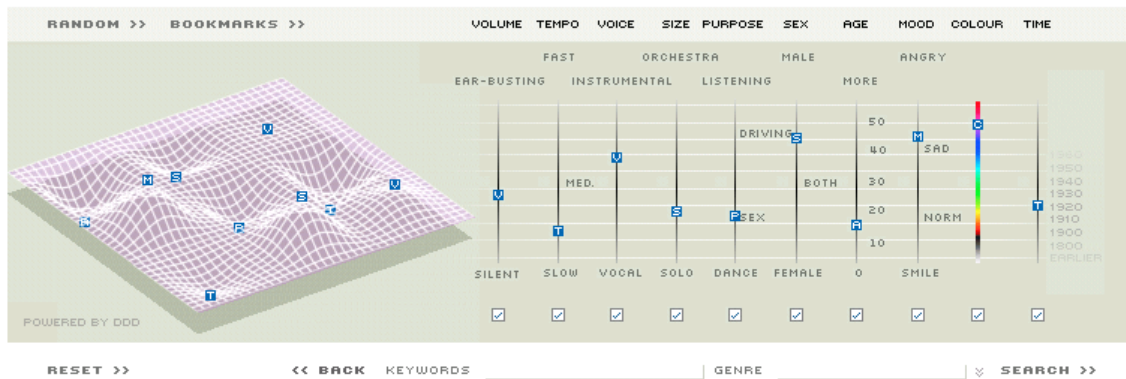


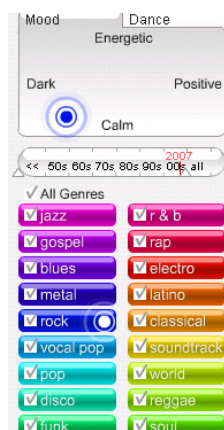
Abbildung 2.4: Künstler und Bands in Beziehung zueinander⁷

Es lässt sich aber genauso gut auf *akustische Metadaten* erweitern. Der Anwender weiß, dass in seiner gesuchten Musik Gitarren vorkommen, dass diese etwas langsamer ist und von der Stimmung her eher melancholisch. Er wählt dafür das Instrument Gitarre aus, in der Temposkala schiebt er den Regler in den Bereich von „eher langsam“ (siehe Abbildung 2.5; wobei Tempo nur eine beispielhafte Eigenschaft ist, die so ausgewählt werden kann) und im Stimmungsfeld navigiert er mehr zum Bereich „melancholisch“ (siehe Abbildung 2.6; z.B. Tanzbarkeit kann auch so ausgewählt werden), d.h. z.B. die Grundtonart ist Moll (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“). Bei der Auswahl wurden hier wieder andere Techniken angewandt (Selektion, Schieberegler und freie Positionierung in einem Feld).

⁷Ausschnitt aufgenommen von <http://audiomap.tunegluemusicmap.net>

Abbildung 2.5: Musikauswahl durch Schieberegler oder 3-D-Netz⁸

In diesem Beispiel hat der Benutzer seinen Ausschnitt auf die Musiksammlung immer weiter verkleinert. Dabei soll zu jedem Zeitpunkt die Möglichkeit bestehen diesen auch wieder zu erweitern oder einfach die Sicht darauf ändern (z.B. von einer räumlichen über eine zeitliche auf eine stimmungsbasierte Sicht). Jede neue Ansicht hat einen gewissen Einfluss auf die Denkweise des Anwenders und die Suchrichtung kann dadurch von mal zu mal geändert werden. Überlagerungen von separaten Kontexten ermöglichen es, erweiterte Zusammenhänge zu erschließen. Wo in Abbildung 2.4 die Zusammenhänge von verwandte Künstler und in Abbildung 2.3 Künstler auf einer Zeitleiste separat dargestellt wurden, werden sie in Abbildung 2.7 übereinandergelegt und noch mit zusätzlichen Informationen angereichert⁹. Diese Variation zeigen nur beispielhaft, dass es viele unterschiedliche Zugangsarten gibt, wie in einer digitalen Musiksammlung navigiert werden kann.

Abbildung 2.6: Auswahl der Stimmung in einem Feld¹⁰

Letztendlich sind die grundlegenden Möglichkeiten im digitalen Bereich deutlich vielfältiger als im physischen Bereich, z.B. die Erfahrung von Orten und Zeit auf den jeweiligen Sichten. Wobei eine reale Bewegung im Raum auch auf eine virtuelle Bewegung in einer Sammlung übertragen werden kann (siehe [SLS08]). Stewart et al.[SLS08] führten

⁸Ausschnitt aufgenommen von <http://finetunes.musiclens.de/>

⁹Die blauen Punkte stehen für einzelne Veröffentlichungen

¹⁰Ausschnitt aufgenommen von <http://musiccovery.com/>

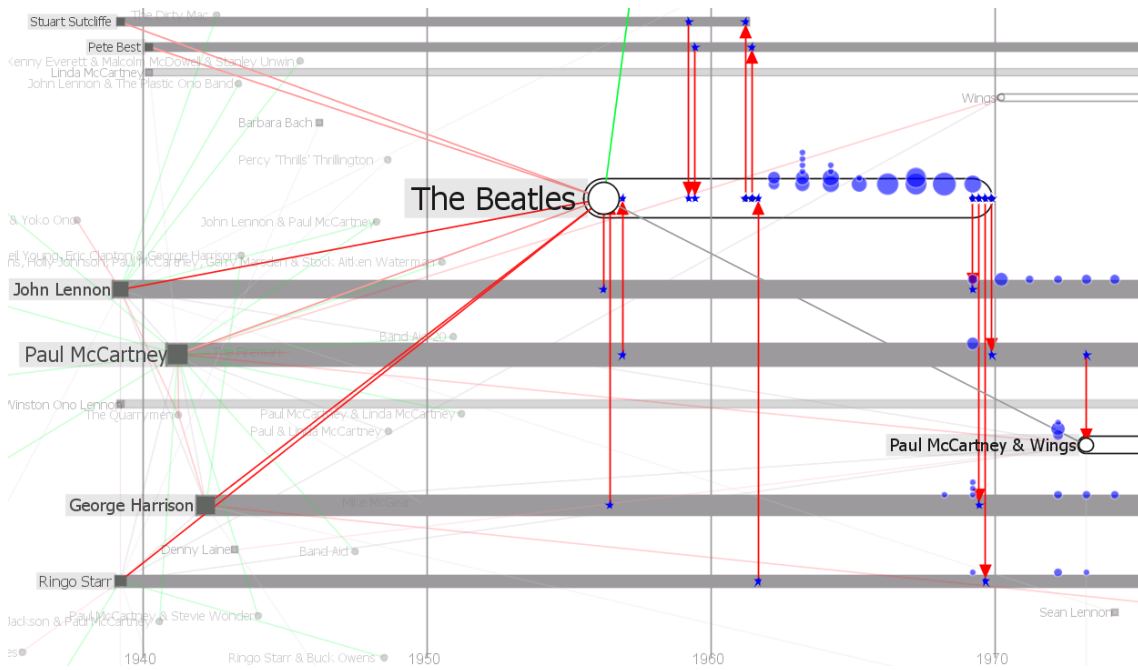


Abbildung 2.7: Künstler, ihre Verbindungen und Veröffentlichungen auf einer Zeitleiste¹¹

dazu ein Experiment durch, bei dem die persönliche Musiksammlung auf einen virtuellen 2-dimensionalen Raum durch Stimmungszuordnungen positioniert¹² wurde und der Benutzer sich zu den Musikstücken im realen Raum hinbewegen bzw. diese heranholen konnte (durch Abbildung der Räume aufeinander).

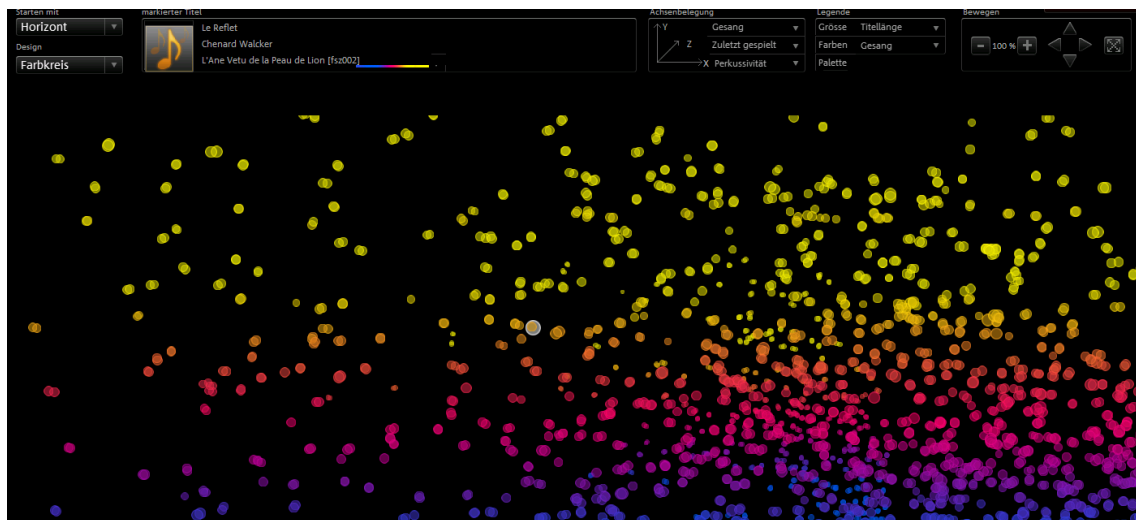


Abbildung 2.8: Räumliche Visualisierung einer Musiksammlung nach den Merkmalen *Gesang, zuletzt gespielt* und *Perkussivität*

Vergleichbar zu einer solchen Anordnung ist die Visualisierung einer Musiksammlung nach festgelegten Eigenschaften in einem virtuellen 3-dimensionalen Raum wie es

¹¹Ausschnitt aufgenommen von <http://www.sfu.ca/~jdyim/musicianMap/musicianMap.html>

¹²nach Aktivität (engl. *activity*; *mild* bis *intensiv*) und Wertigkeit (engl. *valence*; *unangenehm* bis *angenehm*) [SLS08]

z.B. innerhalb der Visualisierungstechnik *mufin vision* realisiert ist (siehe Abbildung 2.8). Hierbei kann sich der Benutzer in der virtuellen Welt frei um die Musikstücke herum bewegen und durch Zuordnung der Merkmale auf Achsen, Größe und Farbe der Exemplare, deren jeweilige Ähnlichkeitsbeziehung erschließen (siehe Abbildung 2.9 und [muf08]). Somit können in dieser Darstellung fünf verschiedene, selbst festgelegte Eigenschaften der Musikstücke auf beliebige Weise zueinander in Relation gesetzt und visualisiert werden. Anita S. Lillie hat sich in ihrer Masterarbeit [Lil08] mit der Navigation innerhalb von persönlichen Musiksammlungen beschäftigt und ist bei der Umsetzung der Visualisierungsarten auf noch weitere interessante Möglichkeiten gestoßen (siehe Abschnitt 2.5). Gerade für diesen Anwendungsfall, wird ein Datenmodell benötigt, welches gut die semantischen Zusammenhänge von verschiedenen Musik- und musikrelevanten Eigenschaften modellieren kann. Die Untersuchungen in Kapitel 4 werden aufzeigen, welches Format dafür geeignet ist.



Abbildung 2.9: Eigenschaftenauswahl bei der räumlichen Visualisierung¹³

2.3 Suchen

Eine Musiksuche soll dem Benutzer auf eine jegliche Form der Anfrage eine möglichst exakte Ergebnismenge liefern. Die dabei genutzten Eigenschaften erstrecken sich teilweise über die gesamte Dimension der *Metadatenfeldergruppierungen* (siehe Abschnitt 4.1 Paragraph „Metadatenfeldergruppierungen“). Nach Tague-Sutcliffe’s Definition ist eine Anfrage, eine in Worte gefasste Aussage einer Benutzernachfrage [TS92]. Da sich Musik nun nicht nur über die mit ihr in Verbindung stehenden *redaktionellen* und *kulturellen Informationen* beschreiben lässt, sind auch Anfragen möglich, welche sich direkt auf den durch das Musiksinal wiedergegebenen (akustischen) Inhalt beziehen. Deshalb lässt sich auch die von Lee et al. [LDJ07] abgeleitete Definition einer *Musikanfrage* wie folgt erweitern:

¹³Ausschnitte von Abbildung 2.8 und 2.9 aufgenommen aus der Anwendung *MAGIX MP3 Maker 15*; siehe <http://www.magix.com/de/mp3-maker/classic/>

Eine *Musikanfrage* ist ein verbalisierter Ausdruck und/oder ein beschreibender Audioinhalt, in welchem Benutzer eine Nachfrage nach einem oder mehreren Musikobjekt(-en) beziehungsweise Informationen über diese ausdrücken.

Somit werden die drei in Cui et al.[CLP⁺07]) beschriebenen *Musikanfragetypen* eingeschlossen. Eine aus Benutzersicht erweiterte Funktionalität bekommen *Anfragen durch Text* (engl. *Query-by-Text*) indem sie hier als *verbalisierte Anfragen* definiert sind (siehe Paragraph „Verbalisierte Anfragen“). Der zweite Typ dieser Kategorisierung, *Anfrage mit Inhalt* (engl. *Query-by-Content*) wurde i.A. übernommen, nur mit der Einschränkung auf Audioinhalt (siehe Paragraph „Anfrage mit Audioinhalt“). Die letzte in diesem Aufsatz beschriebene Kategorie, die *hybride Anfrage*, ist lediglich eine Verschmelzung der vorherig genannten Formen, da diese i.d.R. derzeit getrennt voneinander betrachtet und ausgewertet werden (siehe [CLP⁺07]). Mit diesem Typ kann z.B. die Suche nach einem Musikstück mit einem ähnlich klingenden Beispiel unterlegt werden.

Verbalisierte Anfrage In Worten beschriebene Anfragen im Bereich der Musik unterscheiden sich konzeptuell nicht von denen in anderen Anwendungsbereichen. Da viele existierende Musiksuchmaschinen textuelle Beschreibungen benutzen, um ihre Musikeinheiten zu kennzeichnen [CLP⁺07], ist die *Textanfrage* z.Z. dominierend. Genauso gut und aus Benutzersicht noch intuitiver lassen sich *gesprochene Anfragen* stellen. Zu erst soll hier aber die Grundform beschrieben werden.

Es gibt zwei hauptsächliche Unterscheidungen, die *Stichwortsuche* (d.h. *künstlich*) und die *Suche durch natürliche Sprache* (d.h. Anfragen aus dem wirklichen Leben¹⁴). Aus beiden lassen sich *Suchausdrücke* ableiten, was jeweils eine (aufbereitete) einfache Zeichenkette in der Sprache des Systems (z.B. *SPARQL*, siehe Abschnitt 5.2.2) ist, das die Anfrage auslöst. Ein *Suchprozess* ist somit eine Sequenz von *Suchausdrücken* einer Anfrage [TS92].

Die *Stichwortsuche* kann durch Auswahl von vorgegebenen Termen (stellt die richtige Schreibweise sicher) erfolgen oder durch manuelle Eingabe von Wörtern und kurzen Wortgruppen (lässt mehr Freiheiten bei der Beschreibung), welche auch durch Vorschläge während des Schreibens ergänzt werden kann. Dabei können die einzelnen Teile einer Anfrage durch logische Operatoren miteinander verbunden werden, um die Suchanfrage zu spezifizieren. Zusätzlich können durch die Verwendung von Thesauri Synonyme und Varianten des Suchterms gefunden werden und damit die Ergebnismenge ausgeweitet werden. Durch diese *subjektbasierte Suche* (*subject searching*, siehe [Les06]) können generell alle *Metadatenfelder* und ihre Wertebelegungen abgedeckt werden, d.h. es kann sowohl nach *redaktionellen Informationen*, wie z.B. Künstler und Titel, als auch nach inhaltsbasierten Eigenschaften (Rhythmus, Grundtonart) oder subjektiven Zuordnungen (Stimmung, *Musikkontext*) gesucht werden.

Ein offener Ansatz ist die Eingabe von natürlicher Sprache, d.h. der Anfrager gibt Wortgruppen oder ganze Sätze vor, welche zusätzlich durch *NLP*-Techniken ausgewertet werden müssen. Dies ermöglicht z.B. die Suche mit Ausschnitten aus Liedtexten (siehe [WZO07, GCD07]) oder inhaltlichen Beschreibungen (*Query-by-Semantic-Descriptions*, siehe [TBTL07, CIR⁺08]). Dabei muss die Anfrage zuerst auf Rechtschreibung überprüft werden, damit danach eine inhaltliche Auswertung dieses Terms erfolgen kann. Es ergibt

¹⁴engl. *real-life queries*

sich somit eine Art erweiterte Schlagwortsuche, wobei die Schlagwörter durch Prädikate semantisch miteinander verbunden sind. Ontologiebasierte Metadatenpezifikationen bilden für solche Anfragen die grundlegende Technologie um beschreibende Verbindungen zu erschließen (siehe Abschnitt 4.2).

Alternativ können *verbalisierte Anfragen* auch in gesprochener Form an das System gestellt werden. Da der Benutzer hierzu weder Lese- noch Schreibkenntnisse benötigt, ermöglicht diese Technik einen wesentlich einfacheren Zugang zu einer Musiksammlung als eine Suche durch Texteingabe. Bei der Auswertung einer solchen Anfrage sind weitere Schritte nötig. Dabei werden verschiedene Sprachanalysemechanismen angewandt. Neben der Übersetzung in Text mit Hilfe von Grammatiken usw. kann auch noch ein phonetischer Vergleich des Signals auf Ähnlichkeit erfolgen (siehe [Ter08]). Das Musiksuch- und empfehlungsportal *Midomi*¹⁵ unterstützt gesprochene Anfragen für Musikstücktitel oder Künstlernamen.

Die *schlagwortbasierte Textsuche* ist generell in jeder Anwendung zur Verwaltung von Musiksammlungen vorhanden. In diesen Programmen ist die manuelle Eingabe für Suchanfragen sehr häufig vertreten, welche oft durch eine *ad-hoc*-Aktualisierung der Ergebnismenge(n) während des Tippens ergänzt wird (z.B. *WinAmp*, *iTunes*). Anfragen mit natürlicher Sprache sind noch nicht in den bekannten Anwendungen zur Organisation von persönlichen Musiksammlungen möglich. Verschiedene Untersuchungen zeigen aber auf, dass gerade unscharfe Anfragen bei Musiksammlungen häufig vertreten sind (siehe [SF00]).

Anfrage mit Audioinhalt Bei der Methode *Anfrage mit Audioinhalt* wird immer ein Ausschnitt eines Musikstückes in einer bestimmten Art und Weise wiedergegeben (oft nachgeahmt) und als Eingabe für die Suche genutzt. Diese Technik hilft dem Benutzer, wenn er die *redaktionellen Informationen*, z.B. Künstlername oder Musikstücktitel, nicht genau weiß aber dafür eine Melodie oder einen Rhythmus im Kopf hat (siehe Abschnitt 3.1.2 und [CLP⁺07]). Als Grundlage zur Auswertung einer solchen Suchanfrage werden nur *akustische Eigenschaften* benötigt, die durch die Audiosignalanalyse gewonnen und angereichert wurden (siehe Abschnitt 5.1). Durch Ähnlichkeitsanalyse werden mögliche Ergebnisse berechnet und empfohlen (siehe Abschnitt 5.2.3). Im Folgenden sollen nun kurz verschiedene Methoden mit Beispielanwendungen genannt werden, die im definierten Anwendungsbereich von Bedeutung sind:

- **Anfrage durch Melodie** (*Query-by-Melody*): wenn die Melodie in irgendeiner Weise mündlich, symbolisch oder instrumental¹⁶ wiedergegeben werden kann [Mar08a], z.B.
 - **Anfrage durch Singen** (*Query-by-Singing*): wenn der Liedtext und die Melodie des Musikstückes teilweise bekannt ist, oder
 - **Anfrage durch Summen** (*Query-by-Humming*): wenn nur die Melodie bekannt ist, z.B. *Midomi*
- **Anfrage durch Klopfen** (*Query-by-Tapping*): um den Rhythmus bzw. eine abstrakte Melodie eines Musikstückes wiederzugeben; die Tonhöhe wird dabei nicht

¹⁵<http://www.midomi.com/>

¹⁶Wobei hier Notenkenntnisse vorhanden sein müssen bzw. ein Instrument erlernt werden muss und diese Methoden deshalb für den definierten Anwendungsbereich weniger von Bedeutung sind

berücksichtigt (siehe [JLY01]), z.B. *Melodyhound*¹⁷

- **Anfrage durch Musikbeispiel** (*Query-by-Example*): wenn ein bereits aufgezeichneter Ausschnitt eines Musikstückes zur Verfügung steht (siehe [HL06a]), z.B. *Musicsurfer*¹⁸

Im Bereich der Anwendungen zur Verwaltung persönlicher Musiksammlungen ist bis jetzt noch keine dieser Techniken zum Einsatz gekommen, obwohl sie unter oben genannten Aspekten interessant sind und dem Benutzer einen natürlichen Zugang zu seiner Sammlung bieten. Diese Methoden würden zwar i.A. nicht so oft angewendet werden (vgl. Statistik von Lee et al. über bevorzugte Suchmethoden [LD04]), aber sie löst eine größere Zufriedenheit beim Benutzer aus, wenn die Anfrage zum gewünschten Ergebnis führt. Darüber hinaus ist durch die wachsende Popularität dieser Methoden auf Online-Plattformen (siehe *Midomi*, welches *Anfrage durch Melodie*, *Anfrage durch Musikbeispiel* und auch gesprochene Musikstücktitel und Künstlernamen unterstützt) davon auszugehen, dass diese auch in naher Zukunft in Programmen zur Verwaltung von Musiksammlungen zum Einsatz kommen. Die Systeme zur Melodieerkennung sollten dabei alle möglichen Methoden zur Erzeugung einer Melodie (Liedtext singen, einsilbig singen, summen, pfeifen) unterstützen, da die Benutzer jeweils verschiedene Methoden bevorzugen (siehe [LML04]).

Durch die allgemeine Verbesserung der Verarbeitung der vorgestellten Anfragen, besonders in den Bereichen Mustererkennung und Ähnlichkeitsberechnung (siehe Abschnitt 5.2.3), sind die Schwierigkeiten, welche bis jetzt einen Einsatz dieser Techniken in Anwendungen zur Verwaltung persönlicher Musiksammlungen ausgeschlossen haben, reduziert.

2.4 Empfehlen

Das Empfehlen von Musik unterscheidet sich aus der Sicht des Benutzer in zwei separate Anwendungsfälle. Auf der einen Seite kann ein Benutzer anderen Benutzern der privaten Musiksammlung aber auch Freunden (z.B. per E-Mail) oder Mitglieder einer Online-Gemeinschaft (z.B. *Last.fm*, siehe Abschnitt 4.3.2 Paragraph „Last.fm“) Musikeinheiten (-stücke, -alben, -künstler, etc.) empfehlen (siehe Abschnitt 2.7). Auf der andere Seite kann er selbst von Empfehlungen Gebrauch machen, indem er die in Abschnitt 2.3 beschriebenen Anfragetechniken dazu ausnutzt, um sich Musik empfehlen zu lassen. Diese kann dann in einer geeigneten Repräsentationsform (siehe Abschnitt 2.2) wiedergeben oder ggf. in einer Wiedergabeliste oder sogar einem Mix zusammengesetzt werden (siehe Abschnitt 2.5).

Bei den Empfehlungen kann jede der in Abschnitt 5.2.3 beschriebenen Methoden zur Ähnlichkeitsanalyse verwendet werden. Von Vorteil ist eine ausgewogene Zusammenführung und Auswertung der Ähnlichkeitsmetriken aus:

- *gemeinschaftlichen Filtern*
- *Audiosignalanalyse (inhaltsbasiertes Filtern)*

¹⁷siehe http://melodyhound.com/query_by_tapping.0.html

¹⁸<http://musicsurfer.iua.upf.edu/>

- Metadatenanalyse (inkl. webbasierter Informationsanreicherung- und Verknüpfung, d.h. auch *kontextbasiertes Filtern*)
- Benutzerprofilauswertung

Zwei Charakterisierungen in Bezug auf den Bekanntheitsgrad von *Musikeinheiten*¹⁹ haben i.A. Einfluss auf die Ergebnismenge der Ähnlichkeitsanfrage. Zum einen gibt es die populären (*short-head*) *Musikeinheiten* und zum anderen existieren auch die weniger bis unbekannteren (*long-tail*) Exemplare. Auf *gemeinschaftliches Filtern* basierte Empfehlungen erzielen i.A. schlechte Resultate bei unbekannter Musik, da diese weniger gehört und dadurch oft weniger kommentiert wird (Problem der *Popularitätsverzerrung*). Ähnlichkeitsauswertungen auf reiner Audiosignalanalyse haben diese Popularitätseigenschaft nicht und können somit alle Audiosignalinhalte gleichermaßen behandeln (siehe [Her08]). Des Weiteren unterliegen Benutzerprofile bei der anfänglichen Nutzung eines Empfehlungssystems oft auch dem *Kalt-Start-Problem*, d.h. ohne eine Art *Musikgeschmacksprofil* können nur unzureichend zufriedenstellende Empfehlungen ausgegeben werden. Durch Einbindung existierender *Musikgeschmacksprofile* des Benutzers (z.B. mittels *APML*²⁰) oder Ausnutzung vordefinierter *Ground Truth*-Trainingsdatensätze (Modelle die als Informationsdatengrundlage dienen, siehe Abschnitt 5.2.3), welche auch Stereotypen (d.h. *demografisches Filtern*[Her08]) mit einschließen, kann ein Benutzerprofil grundlegend initialisiert werden. Analog kann das *Kalt-Start-Problem* auch bei neu hinzugefügten Musikeinheiten auftreten, welche anfänglich keine Bewertungen besitzen und somit nicht empfohlen werden (können). Hierbei können, mit Hilfe der durch die Audiosignal- und Metadatenanalyse gewonnenen Daten, auch vorhandene Informationen aus dem Web (z.B. *DBTune*) mit einbezogen werden (siehe Abschnitt 5.2.2).

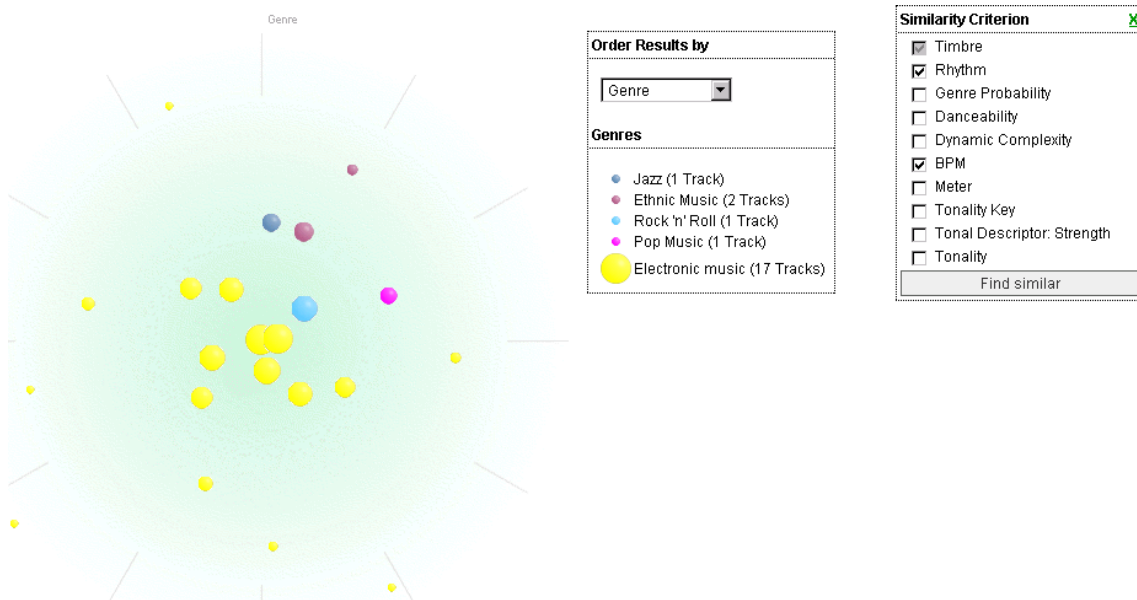


Abbildung 2.10: Auswahl der Merkmale zur Ähnlichkeitsanalyse und Anordnung der Empfehlungsergebnisse²¹

¹⁹Als Sammelbegriff für Musikstücke, -künstler und Zusammenstellungen

²⁰Attention Profile Markup Language; siehe <http://apml.org/>

²¹Ausschnitt aufgenommen von der Anwendung *MusicSurfer*

Nur eine Aggregation bzw. eine sinnvolle Berücksichtigung²² der verschiedenen Ähnlichkeitswerte kann den Problemen der einzelnen Analyseformen entgegenwirken und sorgt somit für eine höhere Benutzerzufriedenheit (vgl. [PR08]). Trotzdem sollten dem Benutzer auch die Auswahl der einbezogenen Techniken und darüber hinaus ggf. die Anpassung von deren Wichtungen möglich sein, da sie eventuell daran interessiert sind, warum das Empfehlungssystem die resultierende Ergebnismenge ausgewählt hat (siehe [Her08]). Die Musikempfehlungsanwendung *MusicSurfer*²³ bietet z.B. die Festlegung der genutzten Musikmerkmale zur Ähnlichkeitsanalyse und auch die Anordnung der Ergebnismenge der Empfehlung (siehe Abbildung 2.10).

Letztendlich können für den Anwendungsfall Musikempfehlung alle *Musikmetadatentypen* (siehe Abschnitt 4.1 Paragraph „Metadatentypen“) verwendet, aber bezogen auf die formulierte Empfehlungsanfrage eventuell unterschiedlich stark genutzt werden. Besonders wichtig im Zusammenhang mit persönlichen Musiksammlungen ist die Berücksichtigung des persönlichen Benutzerprofils mit dem eingeschlossenen *Musikgeschmacksprofil*, welches die Vorgaben zur Auswertung der Metriken der Ähnlichkeitsanalysemethoden vorgibt (siehe Abschnitt 5.2.3).

Abschließend soll für diesen Anwendungsbereich noch die Doktorarbeit von Oscar Celma [Her08] empfohlen werden, welche sich ausführlich mit dem Thema Musikempfehlung und -entdeckung beschäftigt und dabei besonders auf weniger bis unbekannte *Musikeinheiten* eingeht, die bei vielen vorhandenen Musikempfehlungssystemen zu weniger bis gar nicht berücksichtigt werden.

2.5 Wiedergabelisten- und Mixgenerierung

Wiedergabelisten sind eine individuelle oder automatisierte Zusammensetzung von Musikstücken der persönlichen Musiksammlung. Manuelle Aneinanderreihungen von Musikstücken erfolgen i.d.R. nach den Schritten *Orientieren* (siehe Abschnitt 2.2), *Suchen* (siehe Abschnitt 2.3) oder *Empfehlen* (siehe Abschnitt 2.4) durch den Benutzer selbst. Es lassen sich dabei zwei Verwendungsarten der Zusammensetzung unterscheiden (nach [CBF06]):

- **Private Wiedergabeliste:** typischerweise zum persönlichen bzw. nicht formalen Gebrauch, variiert in der Länge und hat i.A. eine weniger strikte Themendefinition
- **Formaler Mix:** gewöhnlicherweise von einer festen Länge (bestimmt durch das genutzte (Ziel-)Medium), hat ein gute definiertes Thema, bei welchem auch die Ordnung von Bedeutung ist

Bei den Erstellungsprozessen soll der Benutzer in jeder Hinsicht intuitiv unterstützt werden. Auf der einen Seite soll er mit möglichst vielen Hintergrundinformationen (z.B. Künstlerbiografien, Albenrezensionen) bei einem Minimum an Interaktion versorgt, um den Vorgaben (ein oder mehrere Themen) gerecht zu werden und geeignete Musikstücke einfacher zu finden. Auf der andere Seite sollen neben präzisen Angaben (z.B. Grundschläge pro Minute, siehe Abschnitt 3.1.2) auch vage Einschränkungen (z.B. „nicht so

²²Ein Verfahren das keine oder unbrauchbare Ergebnisse zurückliefert, sollte nicht mit in die Ergebnismenge fließen; siehe [Her08]

²³<http://musicurfer.iua.upf.edu/>

langsam“) getroffen werden können. Weiterhin soll bei langwierigen Mixerstellungsprozessen die Kommentierung unterstützt werden (z.B. zur Beschreibung einer möglichen Anordnung) [CBF06]. Im Gegensatz zum besonderen Stellenwert eines *formalen Mixes*, laufen *private Wiedergabelisten* oft nur als Hintergrundmusik einer Aktivität. Da besonders bei diesen der Zeitaufwand zur manuellen Erstellung als zu hoch eingestuft wird (siehe [CBF06]), ist ein Forschungsfeld der *MIR*-Wissenschaftler diesen Vorgang zu automatisieren.

Generell soll es einem Anwender möglich sein, auf Grundlage eines jeden *Metadatenfeldes* oder der Kombination dieser, d.h. einem *Musikkontext*, eine Wiedergabeliste von einer Anwendung zur Verwaltung von Musiksammlungen generieren zu lassen. Aus rein technischer Sicht können zwei Hauptformen unterschieden werden: die *durcheinandergemischte*²⁴ *Wiedergabeliste*, welche ohne Nutzereinwirkung entsteht, und die *personalisierte* bzw. *clevere*²⁵ *Zusammensetzung*, die i.A. mehr den Wünschen des Benutzer entsprechen sollte (vgl. [Her08]). Daneben lassen sich im Wesentlichen vier verschiedene Arten der *Wiedergabelistengenerierung* ausgehend von Voraussetzungen bzw. Voreinstellungen unterscheiden:

- **durcheinandergemischt:** d.h. die Auswahl erfolgt auf Basis eines Zufalls- bzw. Mischungsalgorithmus
- **per Anfangswertdefinition** (engl. *seed item*): z.B. Auswahl eines Musikstückes oder Künstler und darauf ausgeführter Ähnlichkeitsanalyse auf Basis von Kontext- und/oder Inhaltsinformationen
- **per Musikkontextdefinition:** und darauf ausgeführter Übereinstimmungsanalyse
- **per Auswahl:** d.h. der Benutzer gibt eine Auswahldimension vor

Bei der *Zufallswiedergabe* müssen die Informationen zu den Musikstücken nicht weiter betrachtet werden. Sie können aber trotzdem hilfreich sein, um den Benutzer wenigstens eine ausgewogene Mischung wiederzugeben, d.h. durch Überprüfung der *Metadatenfeldbelegungen* kann eine Vielfalt bei der Zusammensetzung entstehen. Diese Form ist die einfachste, wobei der Benutzer i.A. überhaupt keine Voreinstellung treffen muss. Deshalb wird sie i.d.R. von den gängigen Wiedergabeanwendungen (*iTunes*, *WinAmp*, *MediaPlayer*, usw.) unterstützt. Selbst diese simple Art kann schon „glückliche Zufälle“²⁶ hervorrufen, z.B. ein Benutzer erhält eine neue Sichtweise auf ein Musikstück, entdeckt unverhofft einen vergessenen Song wieder oder weckt damit alte Erinnerungen (siehe [LVH05]). Generell kann die Durchmischung auch auf jede bestehende Wiedergabeliste angewandt werden und muss dabei nicht einmal ihre eigentliche Anordnung zerstören. Die genutzte Anwendung sollte dem Benutzer dennoch die Möglichkeit bieten, Sequenzen von Musikstücken innerhalb der Zusammensetzung geordnet zu erhalten, damit diese auch wirklich hintereinander abgespielt werden (siehe [CBF06]).

Bei der *Anfangswertdefinition* werden ein oder mehrere konkrete Musikstücke, Künstler oder Zusammenstellungen festgelegt auf deren Basis eine Wiedergabeliste generiert werden soll, d.h. es ist eine Form von *Anfrage mit Audioinhalt* (siehe Abschnitt 2.3). Verschiedene Ansätze können dafür genutzt werden. Durch Analyse von existierenden

²⁴engl. shuffle

²⁵engl. smart

²⁶engl. Serendipity; die Gabe, zufällig glückliche und unerwartete Entdeckungen zu machen

Wiedergabelisten (z.B. von *Last.fm*, siehe Abschnitt 4.3.2 Paragraph „Last.fm“) können weitere Songs herausgefunden werden, die im Zusammenhang mit den gegebenen Informationen stehen. Eine Ähnlichkeit zu diesen muss trotzdem nicht gegeben sein, z.B. kann eine zufällig zusammengestellte Aneinanderreihung oder eine, die für einen bestimmten Anlass erstellt wurde, genutzt worden sein (siehe [Her08]). Auf der anderen Seite, können genauso gut die *redaktionellen* und/oder *akustischen Eigenschaften* der Anfangswertdefinition genutzt werden (wie es z.B. die *mufin audiogen*-Technologie realisiert²⁷). Diese werden i.A. auch für die Übereinstimmungsanalyse auf Basis der *Musikkontextdefinitionen* ausgewertet.

Die dritte Form der *Wiedergabelistengenerierung* stellt ein Grundgerüst für eine veränderbare Wiedergabeliste dar, d.h. aus der Musterdefinition können theoretisch sehr viele unterschiedliche Zusammensetzungen erstellt werden, welche auch dynamisch aktualisiert werden können (siehe [CGD06]). Eine Beschreibung einer solchen Vorlage muss sich dabei aber nicht nur auf positive Beschreibung der in den Musikstücken vorkommende Merkmale beschränken, sondern kann auch negative Bedingungen enthalten, um festzulegen, welche Eigenschaften nicht vorhanden sein sollen (siehe [CBF06]). Bei den letzten beiden Arten sollte unbedingt ein Benutzerprofil existieren, damit die Berechnungen an den Geschmack angepasst werden können. Dies ermöglicht es denn Anwender i.A. auch Rückmeldung zum Gehörten zu geben, z.B. die aktuelle Wiedergabe eines Musikstückes abzubrechen und zum nächsten übergehen²⁸, wenn ihm das Lied nicht gefällt, oder seine Sympathie ausdrücken, wenn er dieses mag (siehe [Her08]).



Abbildung 2.11: Pfadselektion innerhalb einer räumlichen Visualisierung [Lil08]

Die letzte Form der automatisierten Zusammensetzung zielt wieder auf unscharfe Benutzereingaben ab. Mit dieser Art ist nicht die manuelle Auswahl von Musikstücken aus einer textbasierten Liste gemeint, sondern eine grobe Selektion in einer räumlichen Visualisierung der Musiksammlung (vgl. die 3D-Darstellungen im Abschnitt 2.2). In der Masterarbeit von Anita S. Lillie[Lil08] wurden dabei verschiedene Ansätze vorgestellt,

²⁷siehe <http://business.mufin.com/de/produkte/mufin-audiogen-erstellung-von-wiedergabelisten-mit-einem-klick/>

²⁸Wofür das *Metadatenfeld Überspringzähler* geeignet ist

wie in solchen Repräsentationen eine Wiedergabeliste zusammengestellt werden kann. Eine Möglichkeit besteht darin, dass der Benutzer eine Linie innerhalb dieser räumlichen Darstellung (z.B. auch eine Landkarte) zieht und Songs entlang dieses Pfades ausgewählt werden, z.B. beginnend beim Musikgenre Klassik und endend beim Musikgenre Hip-Hop (siehe Abbildung 2.11). Genauso gut lassen sich aber Bereiche auswählen, z.B. Kreis oder Quader, um Musikstücke, in Bezug auf die jeweilige Ähnlichkeitsbeziehungen, als Grundlage zur automatisierten Generierung zu nutzen (siehe Abbildung 2.12).

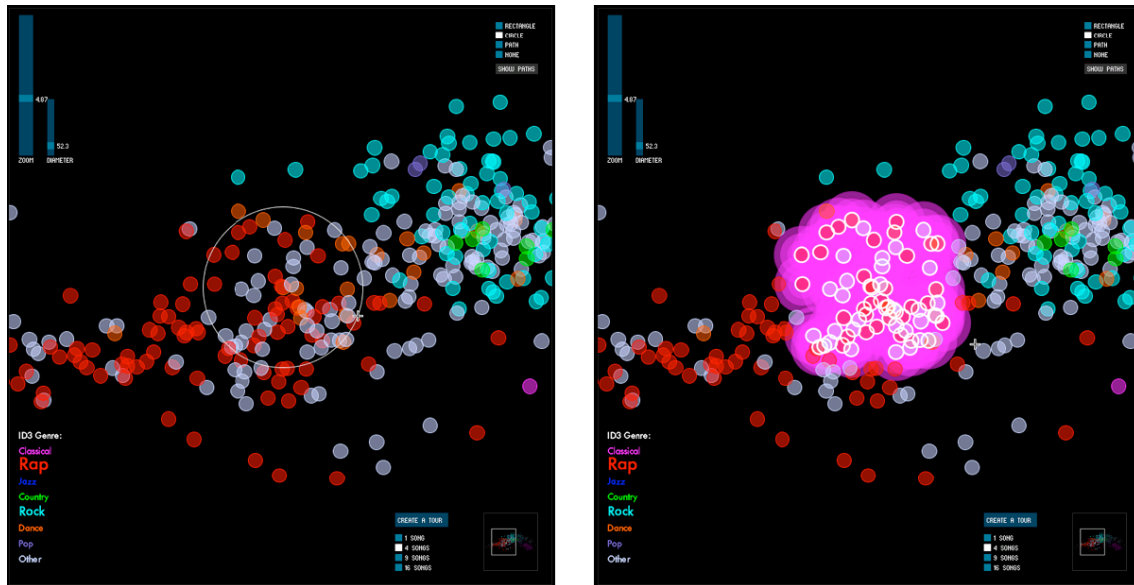


Abbildung 2.12: Kreisselektion innerhalb einer räumlichen Visualisierung [Lil08]

2.6 Verändern

Eine Musiksammlung entsteht durch Veränderung. Am Anfang fängt jeder mit einer Einheit an, egal von welchem Medium sie ist (Musikdokument, CD-Album, Maxi-Schallplatte, usw.). Im Laufe der Zeit wächst eine Musiksammlung in der Regel, d.h.

- es kommen neue Einheiten hinzu (z.B. durch Neuerwerb oder in Form von Sicherungskopien physischer Musikmedien),
- vorhandene Einheiten werden weggenommen (z.B. verkauft, verliehen, gelöscht),
- es ändert sich das Audiosignalträgermedium (z.B. ist eine Schallplatte irgendwann einmal so abgenutzt, dass sie nicht mehr abgespielt werden kann; dann greift der Benutzer auf seine Sicherungskopie in Form eines Musikdokumentes zu),
- oder es ändert sich der (Speicher-)Ort eines Exemplars (d.h. Verschieben oder Kopieren; z.B. Archivieren oder MP3-Player synchronisieren)

Verändern bedeutet aber auch, die Informationen die mit einem Musikstück bzw. einer Einheit in Verbindung stehen, mit neuen Zuordnungen anzureichern oder vorhandene Kategorisierungen und Informationen anzupassen. Dies kann entweder voll- oder teilautomatisiert aber auch manuell erfolgen. Durch die rechnergestützte Verwaltung von Musiksammlungen wird i. A. eine höchst mögliche Automatisierung angestrebt, aber auch ein

einfacher und intuitiver Umgang mit der Menge an definierbaren Eigenschaften. Die (teil-)automatisierte Ergänzung von eingelesenen Audio- und Metadaten erfolgt nach den in Kapitel 5 beschriebenen Techniken und Methoden. Dabei wird versucht die jeweilige Einheit auf Grundlage der vorhandenen Informationen in einen Kontext einzuordnen. Auf der einen Seite sollten durch vorhandene Audiomerkmale bei *akustische Metadaten* eine Zuordnung zu *redaktionellen Eigenschaften* vorgenommen werden können. Auf der anderen Seite können qualitativ kommentierte Musikstücke den bereits abgespeicherten Musikmerkmalen zugeordnet werden. Letztendlich können durch diese beiden *Musikmetadatentypen* auch die *kulturellen Informationen* zugeordnet werden (siehe Abschnitt 4.1 Paragraph „Metadatentypen“).

Nachdem die *Musikanalyseanwendung* eine erste Einordnung vorgenommen hat, kann ein Benutzer jederzeit diese Informationen an seine Verhältnisse anpassen, z.B. eigene Musikgenre- oder Stimmungszuordnungen²⁹, aber auch Korrekturen vornehmen, wenn er der Meinung ist, dass eine Eigenschaft nicht richtig eingeordnet wurde. Dabei sollte er die Möglichkeit besitzen, sowohl einzelne als auch gruppierte Einheiten zu editieren. Diese Abänderungen können ggf. mit einer zentralen Datenbank synchronisiert werden³⁰.

Neben der einfachen Ergänzung bzw. Abänderung auf textueller Basis, d.h. per Freitexteingabe oder Auswahl aus einer vorgegebenen Liste, sollen dem Benutzer andere natürlichere Zuordnungsmethoden helfen. Auf einer Zeitleiste kann durch Setzen oder Verschieben ein neuer Zeitpunkt definiert bzw. ein vorhandener korrigiert werden. Die selben Prozesse (Setzen und Verschieben) können auch im Kontext einer Landkarte vorgenommen werden, nur dass dann der Ort anstelle von der Zeit verändert wird. Generell können demnach alle im Abschnitt 2.2 vorgestellten Sichten zum Verändern genutzt werden.

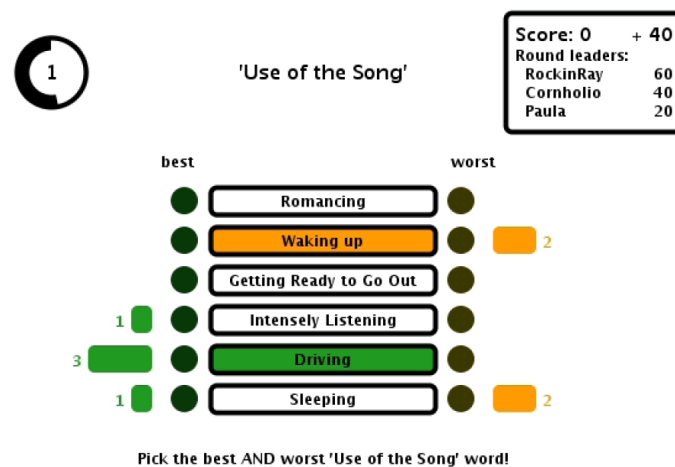


Abbildung 2.13: *Listen Game*: Auswahl der besten und schlechtesten Beschreibung eines Musikstückes [TLBL07]

Eine weitere Möglichkeit hat sich in den letzten Jahren durch diverse *Musikkommentierungsspielideen* im Forschungsbereich hervorgetan. Beispielhaft soll hier nur kurz das *Listen Game* von Turnbull et al.[TLBL07] erwähnt werden, bei welchem mehrere Spieler gleichzeitig über die ihrer Meinung nach beste und schlechteste Beschreibung eines

²⁹Welche i.d.R. parallel zu den durch Gemeinschaften und/oder Experten entstanden existieren und diese ergänzen sollen

³⁰ggf. müssen die Daten zuvor noch zur Wahrung der Qualität validiert werden

Musikstückes aus einer vorgegebenen Auswahl (von Instrumenten, Stimmungen, Genre usw.) entscheiden (siehe Abbildung 2.13) aber auch eigene Kommentierungen in einem weiteren Modus frei eingeben können. Dadurch werden Verbindungen zwischen Schlagwörtern, die in einem strukturiertem Vokabular vorkommen, und Musikstücken gestärkt oder geschwächt und helfen somit, diese besser in ein Gesamtbild einzuordnen. Im Allgemeinen hat dieser Ansatz aber noch nicht den erhoffen Anklang gefunden, da die Spieler häufig dazu neigen, das Spiel auch wirklich zu spielen, d.h. es werden schneller allgemeingültige Beschreibungen, z.B. Rock, ausgewählt anstatt semantisch präzise, z.B. Grunge (siehe [TBL08]).

Nicht zuletzt soll auch das Auslagern und Löschen von Musikedokumenten durch die Anwendung zur Verwaltungen von Musiksammlungen unterstützt werden. Dabei können *Gebrauch-* und *Benutzungsanalysen* dem Benutzer Auskunft über z.B.

- wenig bis gar nicht verwendete Musikstücke,
- Musikstücke, welche in letzter Zeit sehr wenig gehört wurden oder auf eine bestimmte Art nicht gemocht werden (z.B. von einer Wiedergabeliste ausgeschlossen oder oft übersprungen),
- oft oder überhörte Musikstücke (*Überlastung-* und *Erbrechenszyklus* siehe [CJJ04]) oder
- doppelt vorkommende Musikstücke

geben und ihm somit beim Prozess der Verkleinerung der Musiksammlung helfen. Derzeitig bieten die Anwendungen zur Verwaltung von Musiksammlungen teilweise schon Hilfswerkzeuge für diese Aufgabe an und können teil- oder vollautomatisiert ablaufen (z.B. *Songbird's Exorcist*³¹).

2.7 Gemeinsam Benutzen

Eine Musiksammlung wird selten allein genutzt, d.h. Benutzer neigen dazu, Musik, die ihnen gefällt, weiteren potentiellen Interessenten zukommen zu lassen. Generell lassen sich dabei drei Fälle unterscheiden, die im Folgenden kurz erläutert werden.

Der erste Fall handelt davon, dass ein Benutzer Teile seiner persönlichen Musiksammlung in eine gemeinsam genutzte mit einfließen lässt. Zum Beispiel existiert an Orten wie dem Wohnzimmer einer WG oder dem Büro einer kleinen Firma ein Stapel von CDs neben der Stereoanlage zu denen jeder Beteiligte seine Lieblingsstücke dazu beisteuert. Die dabei verloren gegangene persönliche Ordnung (z.B. eine Favoritenrangliste) kann im digitalen Bereich problemlos durch persönliche Benutzerprofile mit angepassten Sichten kompensiert werden (siehe Abschnitt 2.1 und [CJJ04]). Benutzer haben die Möglichkeit, über ein lokales Netzwerk einfach ein gemeinsames Archiv zu pflegen oder Teile ihrer persönlichen Musiksammlung auf dem Rechner freizugeben.

Im zweiten Fall geht es darum, dass ein Benutzer seine persönliche Musiksammlung in irgendeiner Form vorstellt. Zum Beispiel ermöglicht er es einen Interessenten über eine Anwendung zur Verwaltung von Musikarchiven sein eigenes zu durchsuchen. Es wird hierbei gemeinsam Musik gehört und sich über deren Inhalte ausgetauscht. Freunde können

³¹<http://addons.songbirdnest.com/addon/216>

dadurch u.a. neue Musikstile kennen lernen oder sich neue Ordnungsstrategien aneignen (siehe [CJJ04]). Wichtig ist dabei die Möglichkeit der Zugangsbeschränkung, da nicht jeder gern alle Teile seiner Sammlung jeden präsentieren möchte. Programme wie z.B. *MediaMonkey*³² haben diese Funktionalität in Form eines konfigurierbaren „Party-Modus“, bei welchem der verwaltende Administrator die Zugriffsmöglichkeiten auf das persönliche Archiv beschränken kann und somit die eigene Ordnung erhalten bleibt.

Die ersten beiden beschriebenen Fälle führen dabei oft auch zum dritten Fall des *gemeinsam Benutzens*, dem Verleihen. Wo im Physischen das Verleihen von Musikmedien, wie CDs oder Schallplatten sehr oft vermieden wird, existiert es im digitalen Bereich z.Z. eigentlich gar nicht, da Musikdokumente i.A. problemlos kopiert werden können und sich noch keine einheitlichen und wirklich sicheren Kopierschutzmechanismen durchgesetzt haben (siehe [CJJ04]). Durch diese Gegebenheiten wird der natürliche Jäger- und Sammlertrieb des Menschen gerade in dem Bereich von Musikdokumenten verstärkt und die persönliche emotionale und besitzbezogene Verbindung zu einem Audiosignalträgermedium geschwächt. Der wesentlich Grund dabei ist einfach der fehlende physische Aspekt, den eine Musikdokument nicht verkörpern kann.

Eine Folge dieses Verhaltens ist eine ständig wachsende digitale Musiksammlung, die oft in einer Unordnung endet³³, obwohl Kategorisierungstechniken existieren (siehe Abschnitt 5.2.3). Dieser Entwicklung kann durch simples Löschen von Dateien entgegengewirkt werden aber selbst dieser Prozess benötigt einen gewissen Zeitaufwand. Dabei existieren auch schon Hilfsanwendungen, welche den Benutzer beim Aussortieren unterstützen (siehe Abschnitt 2.6).

Um dem natürlichen Jäger- und Sammlertrieb beherrschen zu können, kann z.B. die Methode des *digitalen Ausleihen* (engl. *digital loaning*) eingeführt werden. Hierbei sollte der Prozess des physischen Ausleihens originalgetreu abgebildet werden, d.h.

- es gibt gewisse Laufzeiten (obwohl die beim privaten Ausleihen oft nicht genau festgelegt sind)
- der Ausleihende kann in dieser Zeit nicht auf die Einheit zu greifen (dies kann im digitalen Bereich nur schwer realisiert werden; schon eine einfache vom Rechner entkoppelte Kopie eines Musikdokumentes auf einem dazugehörigen MP3-Player ist eigenständig)
- die am Ausleihprozess beteiligten Personen können miteinander kommunizieren (z.B. die Ausleihdauer verlängern oder den Ausleiher an die Rückgabe erinnern)

Mit Hilfe von Beschreibungsschemata für digitale Rechteverwaltung (sogenannten *Right Expression Languages*; *REL*), z.B. *MPEG-21-REL* (siehe Abschnitt 4.2.2.1 Paragraph „MPEG-21“) oder *ODRL (Open Digital Rights Management)*³⁴, kann dieser Vorgang modelliert werden (siehe Abbildung 2.14). Garcia hat in seiner Doktorarbeit[Gon05] diese beiden *RELs* als Grundlage für die Definition einer allumfassenden *Copyright Ontology*³⁵ genutzt und somit ein semantisches System zur digitalen Rechtsverwaltung geschaffen.

³²<http://www.mediamonkey.com/>

³³z.B. werden die Dateien einfach irgendwo auf der Festplatte abgelegt

³⁴siehe <http://odrl.net>

³⁵<http://rhizomik.net/ontologies/2008/05/copyrightonto.owl>

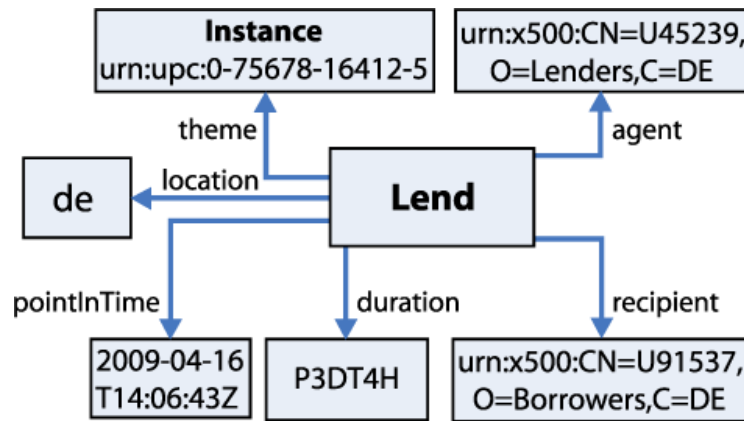


Abbildung 2.14: *Digitales Ausleihen* mit der *Copyright Ontology* modelliert³⁶

Das Anliegen von *digitalem Ausleihen* liegt vorrangig darin der Sinnesüberreizung durch ein Überangebot an vorhandenen Medien entgegenzuwirken und den natürlichen Sammlertrieb auf eine angemessene Ebene zu bewegen. Dadurch wird die Schwelle zur Besitzergreifung auf natürliche Weise hochgeschraubt und ein Benutzer vergrößert seine Sammlung nicht unnötig sondern fügt nur die Einheiten hinzu, welche er auch wirklich haben möchte. Demnach führt er bei frei verfügbaren Exemplaren eine einfache Kopietransaktion durch oder bei kommerziell erhältlichen Musikstücken oder -zusammenstellungen erst einmal eine Transaktion zum Erlangen der benötigten Besitzrechte, d.h. er kauft eine Kopie (vgl. [Gon05]). Kopierschutzbestimmungen spielen beim *digitalen Ausleihen* hingegen eine untergeordnete Rolle, da es sich hierbei mehr um das Ausborgen von digitalen Musikmedien im privaten Umfeld handelt.

2.8 Fazit

Wie die in diesem Kapitel beschriebenen, wesentlichen Anwendungsfälle im Umgang mit Musiksammlungen im privaten Umfeld gezeigt haben, sind die Anforderungen an die Musikmerkmale und -metadaten, die in den folgenden zwei Kapiteln vorgestellt werden, recht komplex und umfangreich. Ein Benutzer möchte in einer einfachen und intuitiven Art und Weise Zugang zu seiner digitalen Musiksammlung bekommen und diese über unterschiedliche und vielfältige Sichtweisen erleben und bearbeiten. Die Audiosignale der Musikdokumente, sollten deshalb recht tiefgründig extrahiert, ausgewertet und zu verständlichen Beschreibungen abstrahiert werden, welche der Benutzer gern anwenden möchte und die auch vage Suchanfragen ermöglichen. Dasselbe gilt auch für die Musik- und musikrelevanten Metadaten, welche in einem reichhaltigen Umfang vorhanden sein und semantisch sinnvoll miteinander verknüpft sein sollten. Der komplexe Prozess der Merkmalsextraktion, der notwendig ist um eine breite *Musikwissenbasis* für die genannten Anwendungsfälle zu erstellen, wird in Kapitel 5 beschrieben.

³⁶Beispiel ist von der *Rent*-Aktion abgeleitet; siehe <http://rhizomik.net/~roberto/thesis/html/Conceptualisation.html#ActionRent>

3 Musikmerkmale

Das komplexe Phänomen Musik lässt sich in seinem Umfang durch eine Vielzahl von definierten Konzepten aber auch subjektiv empfundenen Eigenschaften beschreiben, welche als Musikmerkmale zusammengefasst werden. Dabei haben sie ihren Ursprung in der Physik, Psychoakustik, Musiktheorie und Musikwissenschaft, oder sind einfach von einer Gesellschaft oder Kultur geformt worden. Die Musikmerkmale sind deshalb sehr vielseitig und lassen sich auf unterschiedlichen Ebenen und Unterebenen anordnen. Angelehnt an das konzeptuelle Rahmenwerk *MAMI* von M. Lesaffre (siehe [Les06]) ergibt sich eine Unterteilung in verschiedene Kategorien mit Unterbereichen wie in Abbildung 3.1 zu sehen ist und welche die Unterkapitel dieses Abschnittes bilden.

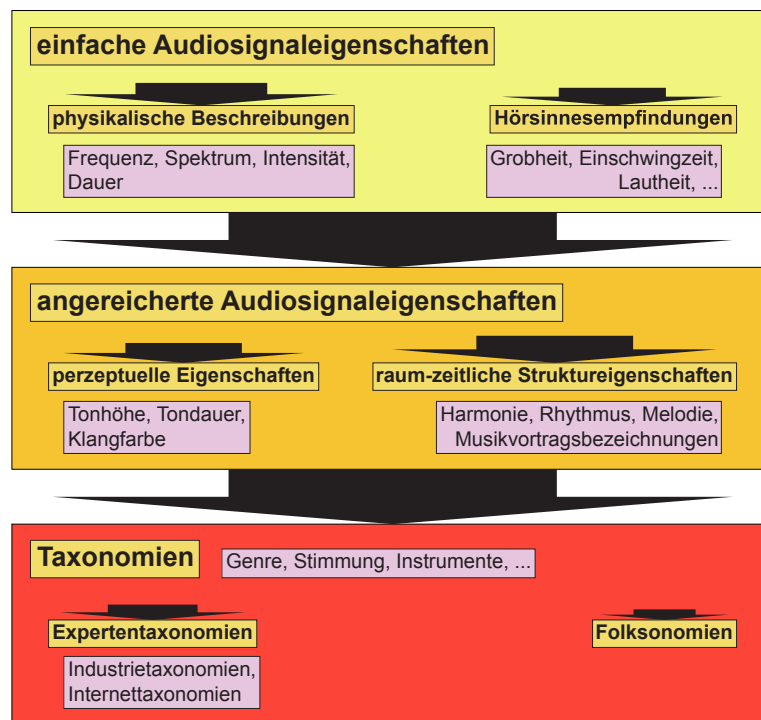


Abbildung 3.1: Kategorien von Musikmerkmalen mit Unterbereichen und Beispielen (flieder-farben)

Die Grundlage zum Empfinden, Wahrnehmen und Bewusstwerden von Musik mit ihren umfassenden Merkmalen zur Beschreibung bildet das *auditorische System*. Dieses lässt sich grob in zwei Bereiche aufteilen. Zu allererst werden durch Objektschwingungen entstandene Schallwellen¹ in Form von Luftdruckschwankungen am *peripheren auditorischen System* (siehe Abbildung 3.2) empfangen, welche Membranschwingungen im Ohr

¹*Schall* als Schwankungen des Luftdrucks im hörbaren Frequenzbereich

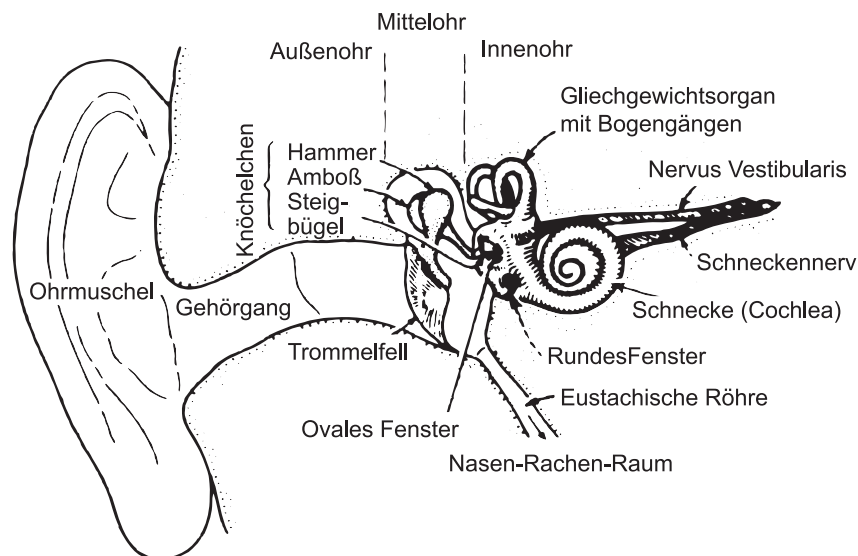
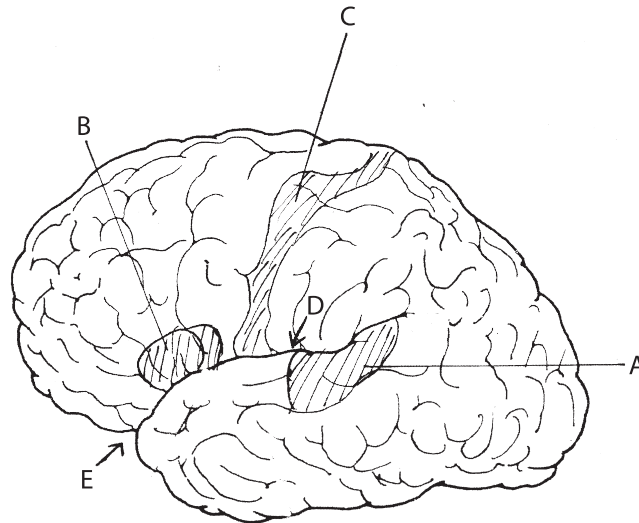


Abbildung 3.2: Schema des peripheren auditorischen Systems [Roe00]

auslösen (siehe Unterkapitel 3.1.1.2). Dabei werden die mechanisch ankommenden Signale in electrophysiologische Reize umgewandelt und über Hörnerven an den *auditorischen Kortex* weitergeleitet, welcher auch als *primäres akustisches Rindenfeld* bezeichnet wird und Teil des Temporallappens im Gehirn ist.

Jede Gehirnhälfte besitzt einen solchen auditorischen Kortex, wobei es aber immer einen dominanten Kortex (oft auf der linken Hemisphäre) gibt an den die Reize des anderen weitergeleitet werden. Von dort aus werden die akustischen Reize dann in den Assoziationsarealen (*sekundärer auditorischer Kortex*) weiter verarbeitet wo ein Abgleich mit anderen Sinnesmodalitäten (z.B. visuelle Wahrnehmungen) stattfindet und evtl. kortikale Reaktionen hervorruft (z.B. motorische Antworten) [dlMHR05]. Deshalb gibt es auch kein isoliertes Zentrum im Gehirn, das für die Musikwahrnehmung verantwortlich ist, sondern ein funktionelles und plastisches (räumliches) Netzwerk verschiedener Hirnareale, die auf jeweilige musikspezifische Erfordernisse spezialisiert sind (z.B. Tonhöhenanalyse, Intervallanalyse, Rhythmusanalyse, siehe Unterkapitel 3.1.2 und 3.2). In Abbildung 3.3 sind diese einzelnen angesprochenen Bereiche des Gehirns markiert und kurz zur Einordnung erläutert.

Eine gute Zusammenfassung des komplexen Prozesses der Musikwahrnehmung und -verarbeitung durch den Menschen – der auch *auditorische Szenenanalyse* genannt wird (engl. *Auditory Scene Analysis*; ASA [Mar08b]) – ist in Abbildung 3.4 illustriert. Die Darstellung konzentriert sich dabei auf die höherwertige Verarbeitung der Signale durch das Gehirn und nicht auf die Hörsinnesempfindungen, welche direkt im Ohr auftauchen (siehe Abschnitt 3.1.1.2). Dabei werden Erste raum-zeitliche Gruppierungen in Klangergebnisse durch *auditiven Gruppierungsprozesse* vorgenommen, wovon *perzeptuellen Eigenschaften* abgeleitet werden können. Diese können danach oder später durch *abstrakte Wissensstrukturen* interpretiert, identifiziert und eingeordnet werden. Wobei die gewonnenen Zusammenhänge bei nachfolgenden Ereignissen mit Einfluss nehmen können, welche durch die *Ereignisstrukturverarbeitung* analysiert werden. Somit entsteht nach und nach eine *gedankliche Darstellung der Schallumgebung*. Diese Vernetzung kann letztendlich auch noch die *Aufmerksamkeitsprozesse* beeinflussen und Erwartungen und Hoffnungen erzeugen, steigern oder abschwächen (siehe [Mar08b]).



- **A:** auditorischer Kortex
- **B:** motorisches Sprachzentrum
- **C:** Gyrus postcentralis (mit dem somatosensiblen Kortex für die Verarbeitung der taktilen Wahrnehmung)
- **D:** Planum temporale (als Teil des Temporallappens mit einem auditorischen Assoziationskortex)
- **E:** limbisches System (für die affektive Bewertung der Reize und wesentlicher Vorgänge des Gedächtnisabgleiches, d.h. zur Erkennung von Musik als solche)

Abbildung 3.3: Für die Musikwahrnehmung relevante Hirnareale auf der linken Hirnhälfte (mit Begriffslegende) [dlMHR05]

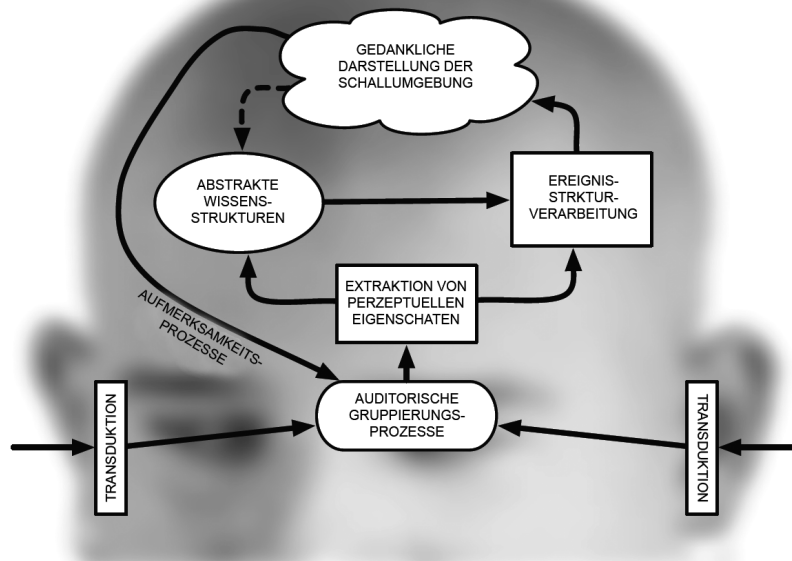


Abbildung 3.4: Die wesentlichen auditorischen Verarbeitungsblöcke der ASA (nach [Mar08b])

3.1 Audiosignaleigenschaften

Audiosignaleigenschaften stellen alle messbaren, empfindbaren, perzeptuellen und strukturellen Merkmale eines Audiosignales dar, welches in Form von Schallwellen durch das menschliche Hörsystem (*auditorische System*) aufgenommen und anschließend vom Gehirn interpretiert und eingeordnet wird.

Zuerst wird im folgenden Abschnitt auf die einfachen, grundlegenden Beschreibungen eines Audiosignales eingegangen (siehe Abschnitt 3.1.1), um darauf aufbauend danach die angereicherten Audiosignaleigenschaften vorzustellen und zu erläutern (siehe Abschnitt 3.1.2).

3.1.1 Einfache Audiosignaleigenschaften

Die einfachen Audiosignaleigenschaften, welche auch als *lokale* bzw. *nicht-kontextuelle Deskriptoren* bezeichnet werden (siehe [Les06]), charakterisieren quantifizierbare Klangeigenschaften. Sie unterteilen sich dabei in zwei grundlegende Merkmalskategorien. Auf der einen Seite befinden sich die physikalischen Beschreibungen eines Signals, wie Frequenz, Dauer, Intensität oder Spektrum. Auf der anderen Seite werden die Sinnesempfindungen des menschlichen Hörsystems erfasst, wie Grobheit, Einschwingzeit oder Lautheit.

3.1.1.1 Physikalische Beschreibungen

Die physikalischen Beschreibungen eines Audiosignals werden auch als *akustische Merkmale* bezeichnet (vgl. [Les06]), wobei die in dieser Kategorie bestimmten Eigenschaften als widerspruchsfrei bzw. objektiv gelten. Im Folgenden werden nun wesentliche Maße und ihre Bedeutung in der Musik kurz erläutert.

Zeit und Dauer Der zeitliche Aspekt innerhalb der Musikmerkmale wird von verschiedenen absoluten und relativen Zeitangaben der physikalischen Größe *Zeit* (t) geprägt. Somit gibt es Zeitpunkte (Ereignisse; z.B. „ab Sekunde 30 des Musikstückes setzt der Gesang ein“) und Zeitintervalle (*Dauer* oder *Zeitraum*; z.B. „der Refrain dauert 20 Sekunden“ oder Musikstückdauer), welche zur Beschreibung genutzt werden können. Wenn sich die Musikmerkmale direkt auf die Auswertung des Audiosignals beziehen, sind die Einheiten *Millisekunde* (ms), *Sekunde* (s), *Minute* (min) und *Stunde* (h) die häufigen Bezugsmaße (Zeiträume). Im Gegensatz zu dieser exakten Analyse, werden kontextuelle Informationen, wie z.B. Entstehungs- oder Veröffentlichungszeitpunkt, oft nur ungenau in Form des *Jahres* (a) angegeben.

Segmentierungen in der Musik, wie z.B. Noten, Kehrreim oder Strophe, werden oft durch diskrete Zeitangaben beschrieben (siehe Abschnitt 3.1.2.1 Paragraph „Tondauer“ und Abschnitt 3.1.2.2 Paragraph „Rhythmus“). Sie können aber bei der Interpretation dieser Klangereignisse, d.h. die Darbietung eines Musikstückes, in absolute und relative Zeitangaben der kontinuierlichen Dauer gemessen und berechnet werden.

Frequenz Grundlegend besitzt jeder einzelne Ton eine *Frequenz*. Dies ist in diesem Fall die Anzahl der periodischen Schwingungen einer Schallwelle pro Sekunde, welche

auf unser Ohr (peripheres auditorisches System) auftrifft und in eine Membranschwingung umgewandelt wird. Mit dieser *elastischen Welle* wird *akustische Energie* (das Produkt aus Intensität Dauer) fortgepflanzt, welche benötigt wird, um jeden von der Welle erreichten Punkt in Bewegung zu versetzen² (siehe Paragraph „Intensität“). Die Einheit der Frequenz ist *Hertz (Hz)* und gibt die Schwingungen pro Sekunde an. Sie bildet somit die Basis für den räumlichen Aspekt innerhalb der Musikmerkmale.

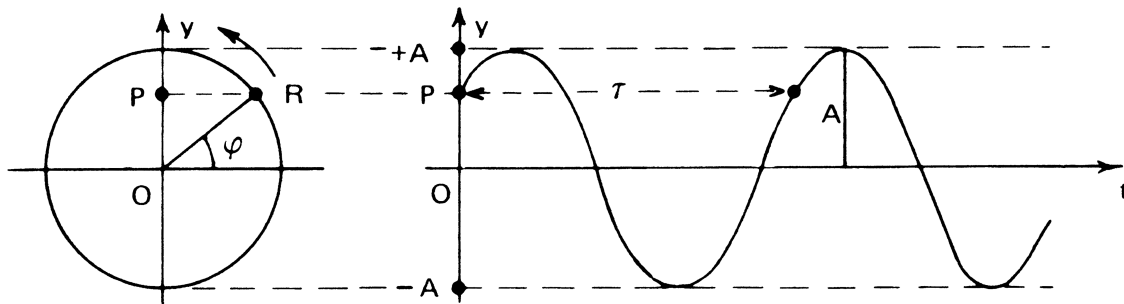


Abbildung 3.5: Projektion eines Punktes in gleichförmiger Kreisbewegung [Roe00]

Das einfachste Modell einer Schwingung, ist die *Sinusschwingung* (auch *einfache harmonische Schwingung*), welche ein Abbild der Sinusfunktion als Projektion einer Kreisbewegung ist [Hem01]. Während des Ablaufs einer harmonischen Schwingung wird ständig potentielle Energie in kinetische und wieder zurück umgewandelt (siehe Paragraph „Intensität“ und [Roe00]). Abbildung 3.5 zeigt ein Beispiel einer Sinusfunktion, welche auch die Aufzeichnung eines *reinen Tons* sein könnte. Durch *Reibung* nimmt nun die mechanische Gesamtenergie ständig ab und vermindert die Amplitude, so dass eine *gedämpfte Schwingung* entsteht.

Die Zeit für eine Kreisbewegung (Periode) gibt die *Periodendauer* τ an und die größte Auslenkung dieser Projektion nennt man *Amplitude* A . Darüber hinaus gibt es noch die *Phase* ϕ , welche den Winkel zum Anfangszeitpunkt $t = 0$ beinhaltet. Dieser Parameter gewinnt erst an Bedeutung, wenn sich Frequenzen überlagern und die *Phasendifferenz* als Unterschied zwischen zwei Phasen mit in Betracht gezogen wird. Dies ist in natürlich wahrnehmbaren Klängen (als zusammengesetzte komplexe Schwingungen) oft der Fall. Durchschnittlich hat der Mensch je nach Alter und Empfindungsvermögen einen Hörbereich von ca. 20 Hz bis 15–20 kHz (siehe [Roe00] und [Hem01]). Der darin befindliche maßgebende Teil für Musik befindet sich zwischen 20–4000 Hz³ (siehe Abb. 3.12 und [Roe00]). Ein Ton der Tonhöhe A' (siehe Abschnitt 3.1.2.1, Paragraph „Tonhöhe“) liegt bei ca. 440 Hz, was auch in der Normierung *ISO 16:1975* festgehalten wurde.

Intensität Die *Intensität* (I) ist der *Energiefluß* oder die Amplitude der Druckschwingungen (auch *Schalldruck* p genannt) der Schallwelle, die das Ohr erreicht [Roe00]. Sie wird deshalb auch manchmal *Schallintensität* genannt und ist als der Betrag der gesamten *mechanischen Energie* (auch *Schallenergie* W genannt, zusammengesetzt aus *potentiell-*

²In diesem Fall der Basilarmembran des Innenohres; siehe Abschnitt 3.1.1.2

³Was einen Umfang von grob 8 Oktaven (siehe Abschnitt 3.1.2.1, Paragraph „Tonhöhe“) ist; manchmal wird der Bereich erweitert bzw. um grob eine Oktave verschoben angegeben (50–8000 Hz), siehe [Par07]

ler⁴ und *kinetischer Energie*⁵; in *Joule*), der pro Sekunde durch eine Einheitsfläche (1 m^2) fließt, die senkrecht zur Ausbreitungsrichtung steht. Der Energiefluß wird in *Joule pro m²* oder *s* angeben bzw. in *Watt/m²* wenn die Definition der Leistung benutzt wird. Dabei existiert ein Zusammenhang zwischen der Intensität und dem *Effektivwert*⁶ der *Druckschwingungen* Δp (in *Newton/m²* bzw. *Pascal*).

$$I = \frac{(\Delta p)^2}{V \cdot \delta} \quad (3.1)$$

Hierbei ist V die Geschwindigkeit der Schallwelle und δ die Luftdichte. Bei normalen Temperatur- und Druckbedingungen ergibt dies ein Verhältnis von:

$$I = 0,000234 \cdot (\Delta p)^2 \text{ (Watt/m}^2\text{)} \quad (3.2)$$

Der leiseste noch wahrnehmbare 1000-Hz-Ton, auch *Hörschwelle* genannt (siehe Abschnitt 3.1.1.2 Paragraph „Lautheit“), hat eine Intensität von nur $10^{-12} \text{ Watt/m}^2$ bzw. eine durchschnittliche Druckänderung von $2 \cdot 10^{-5} \text{ Newton/m}^2$, was dem ca. $2,0 \cdot 10^{-10}$ -fachen des normalen atmosphärischen Drucks entspricht. Bei der Schmerzgrenze (siehe Abschnitt 3.1.1.2 Paragraph „Lautheit“) liegt die Schallintensität bei ca. 1 Watt/m^2 . Diese beiden Intensitäten haben also ein Verhältnis von ca. 1.000.000.000.000 : 1, welches deutlich die Empfindlichkeit des menschlichen Ohres ausdrückt (siehe [Roe00]).

Zur besseren Beschreibung der Schallintensität wird der *dekadische Logarithmus* auf das Verhältnis zu einer Bezugsintensität I_0 angewandt und als *Schall-Intensitätspegel* bzw. *Schalldruckpegel* L bezeichnet, wobei I_0 bzw. Δp_0 sich auf die Werte der Hörschwelle bei einer Frequenz von 1000 Hz beziehen.

$$L = 10 \cdot \log \frac{I}{I_0} \text{ (dB)} \quad L = 20 \cdot \log \frac{\Delta p}{\Delta p_0} \text{ (dB)} \quad (3.3)$$

Die Einheit des Schallintensitätspegels ist *Dezibel (dB)* und immer ein relatives Maß auf einen festen Bezugswert, z.B. 0 dB bei der Hörschwelle. Somit hängt das Maß der Lautstärke immer relativ mit den Intensitäten der einzelnen Töne zusammen, die in einem Klang oder Geräusch überlagert werden (siehe Abschnitt 3.1.1.2 Paragraph „Lautheit“). Von der Intensität können weitere statistische Werte abgeleitet werden, welche im Abschnitt 5.1.4.1 näher beschrieben werden.

Spektrum So bald sich Töne verschiedener Frequenzen zu einem *komplexen Ton* (auch Klang) überlagern, ergibt sich von diesem Klang ein *Frequenzspektrum*. Demnach kann man auch jede noch so komplexe, periodische Schwingung (siehe Abbildung 3.6) als die Überlagerung reiner harmonischer Schwingungen darstellen (dem *Satz von Fourier* folgend, [Roe00]). Die Bestimmung der einzelnen harmonischen Komponenten ist mit Hilfe einer *Fourier-Analyse*⁷, hier im Speziellen *Klanganalyse*, möglich und der umgekehrte Prozess heißt *Fourier-Synthese* bzw. *Klangsynthese* (siehe Abschnitt 5.1.2). Die *Grundfrequenz* ist hierbei die Wiederholungsfrequenz der komplexen periodischen Schwingung (siehe auch Abschnitt 3.1.2.1 Paragraph „Tonhöhe“).

⁴Lageenergie

⁵Bewegungsenergie

⁶engl. *root mean square, RMS*

⁷Auch die *Wavelet-Transformation* anwendbar siehe Kap. 5.1.2 bzw. [Lid06]

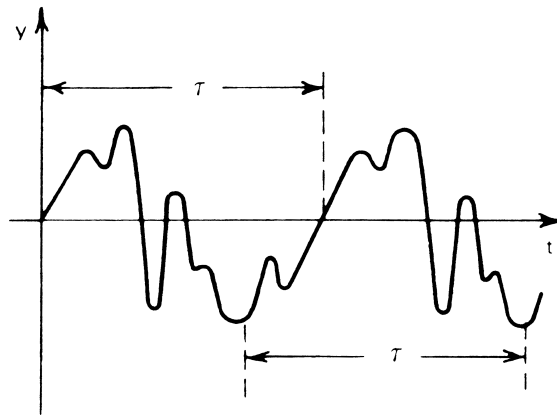


Abbildung 3.6: Eine komplexe periodische Schwingung

Neben dem Frequenzspektrum kann auch noch das *Leistungsspektrum* (auch *Tonspektrum* genannt, siehe [Roe00]) vom komplexen Tönen bestimmt werden. Dies ist die Reihenfolge der Intensitätswerte (siehe Paragraph „Intensität“) der Harmonischen des Klanges. Es trägt im Wesentlichen mit zur Unterscheidung der Klangfarbe bei (siehe Abschnitt 3.1.2.1 Paragraph „Klangfarbe“). Die Tonspektren werden zu den einzelnen Teilfrequenzen aber auch an den Frequenzgruppen (siehe nächstes Unterkapitel) zusammengefasst dargestellt. Neben der Angabe der Intensität wird auch das verwandte Maß des Schalldruckpegels verwendet (siehe Paragraph „Intensität“). Beide Größen werden dabei i.A. relativ zur Grundfrequenz bzw. Gesamtintensität angegeben.

Eine Auswertung bzw. Darstellung eines Tonspektrums in Bezug auf die Zeit nennt man *Spektrogramm* (siehe Abbildung 3.18). Es setzt sich aus den drei Dimensionen Frequenz, Intensität und Zeit zusammen, so dass auch Bereiche wie Ankling- oder Abklingphase mit in die Betrachtung einbezogen werden können (siehe Abschnitt 3.1.1.2 Paragraph „Einschwingzeit“). Daraus lassen sich einige weitere Größen bzw. Statistiken ableiten, welche im Abschnitt 5.1.4.2 näher beschrieben werden. Gerade die umfangreiche Auswertung des Spektrums von Audiosignalen eröffnet die Möglichkeit auf viele höherwertige Musikmerkmale, wie z.B. Tonhöhe oder Rhythmus, zu schließen (siehe Kapitel 5).

3.1.1.2 Hörsinnesempfindungen

Die Eigenschaften der Hörsinnesempfindungen (*auditive Wahrnehmung* [dlMHR05]) ergeben sich aus der ersten Signalverarbeitung des auditorischen Systems und werden auch *sensorische Merkmale* genannt (vgl. [Les06]). Sie beziehen sich auf wahrnehmbare Charakterisierungen von Audiosignalen und Prozesse des *echoischen Gedächtnisses*⁸. Beschreibungen dieser Kategorie können mehrdeutig sein, d.h. sie werden insbesondere auch oft unterschiedlich von verschiedenen Menschen wahrgenommen.

Bei der Verarbeitung von Signalen im menschlichen Hörsystem spielen die *Frequenzgruppen* (auch *Frequenzbänder*; engl. *critical bands*) eine wesentliche Rolle. Nachdem die ankommenden Luftdruckschwankungen durch das Außenohr gefiltert (Reflektionen, Verstärkung der Frequenzen bei 4 kHz, usw.) wurden, werden im Bereich des Mittelohres die Schallwellen von der Luft in eine Flüssigkeit des Innenohres übertragen, wo-

⁸Als akustischer Teil des sensorischen Gedächtnisses (Ultrakurzzeitgedächtnisses) mit der Aufrechterhaltung von Information von bis zu 2s [Nie01]

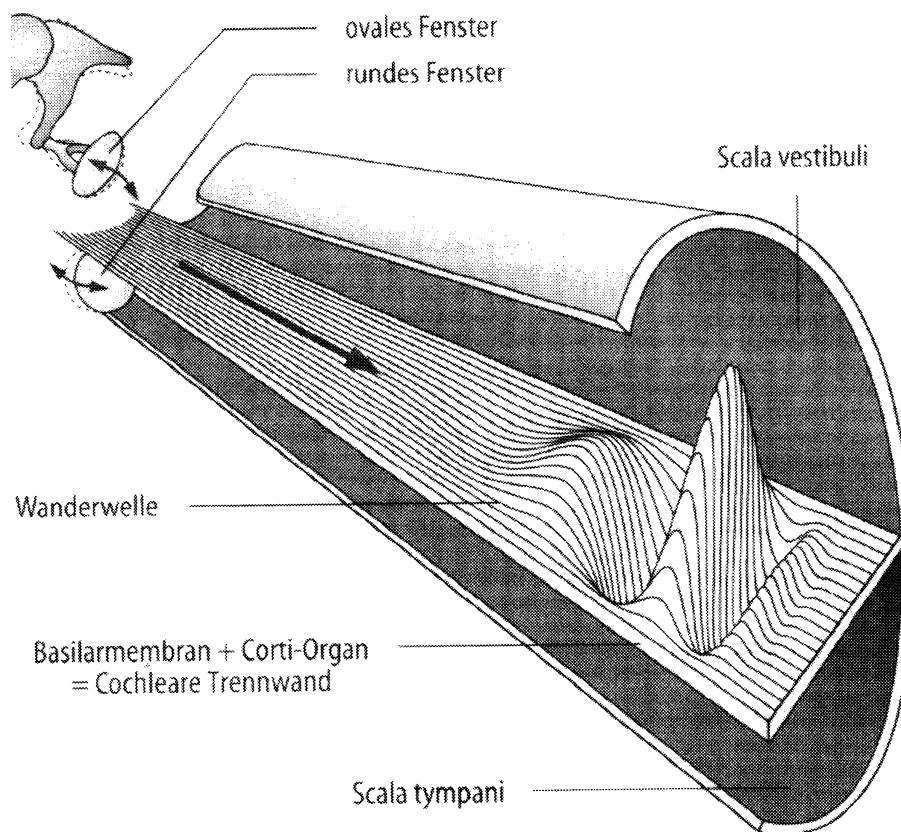


Abbildung 3.7: Wanderwelle entlang der (ausgerollten) Basilarmembran [dIMHR05]

bei noch einmal eine geringe Schwächung des Signals stattfindet (siehe [Jeh05]). Somit wird der Bereich der wahrgenommenen Frequenzen eingeschränkt. Durch einen Hörfrequenzanalysemechanismus werden dann die Eingangssignale mit Hilfe mechanischer Filter in Frequenzgruppen mit bestimmten Bandbreiten zerlegt und somit gemeinsam ausgewertet. Diese Filter sind feine Haarzellen entlang des *Basilarmembrans* in der *Cochlea* (Hörschnecke) des Innenohres, welche über Hörnerven mit dem zentralen Nervensystem verbunden sind (siehe [dIMHR05]). Für den menschlichen Hörbereich ergibt sich eine Unterteilung in 24⁹ Frequenzgruppen, die sogenannte *Bark-Skala*¹⁰ (siehe [FZ07]).

In Abbildung 3.7 ist eine (stehende) Wanderwelle zu sehen, welche durch Druckschwankungen in Folge eines Schallsignales entstanden ist. Die maximalen Amplituden in diesem Beispiel befinden sich recht weit hinten auf der Basilarmembran (hin zur Schnecken spitze) und geben demnach eher eine niedrige Frequenz wieder. Hohe Frequenzen hingegen werden schon an der *Basis*, dem *ovalen Fenster*, aufgenommen. Der bedeutende Frequenzbereich für die Musik (ungefähr 20–4000 Hz) nimmt dabei ca. zwei Drittel der Ausdehnung der Basilarmembran in Anspruch (siehe Abbildung 3.8) [Roe00].

Im Folgenden werden nun die fünf auditiven Sinneswahrnehmungen erklärt, welche für die Audiosignalanalyse von Bedeutung sind.

⁹Eigentlich 25, da die Werte der Bark-Skala von 0-24 gehen und immer jeweils die Zwischenräume ein Frequenzband ergeben

¹⁰benannt nach Heinrich Barkhausen

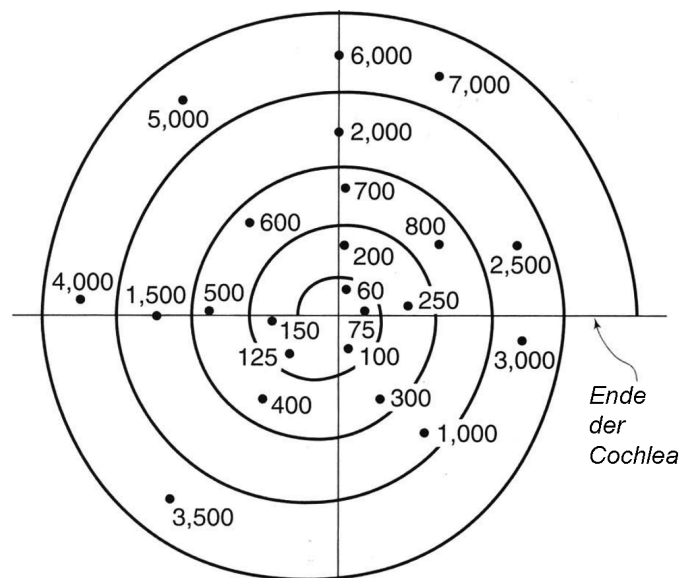


Abbildung 3.8: Ungefähre Frequenzbereiche entlang der Basilarmembran der Cochlea [Gol05]

Tonheit Die Empfindung der Tonhöhe (als angereicherte Audiosignaleigenschaft, siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“), wird als *Tonheit* bezeichnet (siehe [Ter98]). Zur Ermittlung dieser Größe wird zum einen die bereits eingeführte gleichmäßige Unterteilung der Basilarmembran in 24 Frequenzgruppen (*Bark-Skala*) und dem Maß *Bark* benutzt. Dabei entspricht 1 Bark ca. 1,3 mm auf einer normierten 32 mm langen Basilarmembran, was ca. 27 wahrnehmbaren unterscheidbaren Tonhöhenstufen entspricht (siehe [FZ07]).

Ein weiteres Maß zur Tonheitangabe beruht auf der Distanzschätzung von Tonhöhen (z.B. *halbe* oder *doppelte Tonhöhe*) und wurde von Zwicker et al. auch auf die *Bark-Skala* abgebildet. Grundlage war die abweichende Tonhöhenwahrnehmung von halben Tonhöhen bei höheren Frequenzen. Bei niedrigen Frequenzen liegt diese bei einem Frequenzverhältnis von ca. 2:1 (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“). Dahingegen wurde z.B. die halbe Tonhöhe von einem 8 kHz-Ton bei ca. 1300 Hz wahrgenommen, d.h. nicht bei den angenommenen 4 kHz. Die sogenannte *Mel-Skala* mit ihrer Einheit *mel* bringt diese Distanzschätzung nun in ein lineares Verhältnis. So ergibt z.B. ein 8 kHz-Ton eine Tonheit von 2100 mel und die zugehörige halbe Tonhöhe von 1300 Hz einen Wert von 1050 mel.

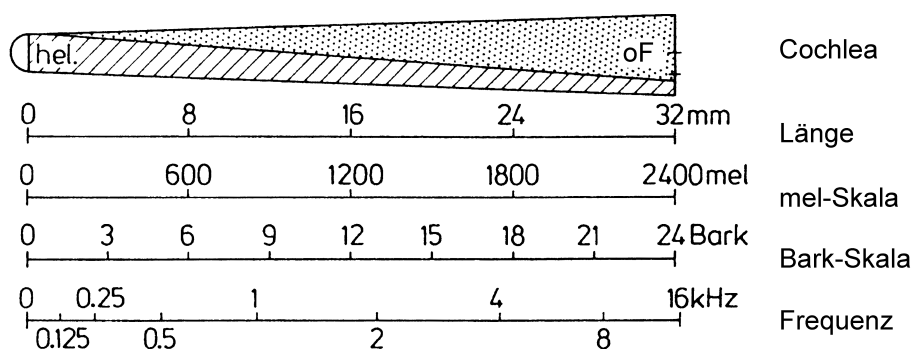


Abbildung 3.9: Die ausgerollte Cochlea mit dem Bezug zu den Tonheitsgrößen Mel und Bark und zur Frequenz

Die mel-Skala verfeinert dabei auch die Unterteilung der Bark-Skala, denn 1 Bark entsprechen grob 100 mel (siehe [FZ07]). Auf der Basilarmembran entsprechen 2 mel dem Abstand von 3 Haarzellen (siehe [dlMHR05]). In Abbildung 3.9 werden diese Zusammenhänge auf einer ausgerollten (normierten) Basilarmembran noch einmal recht gut miteinander verglichen. Wobei die Mel- und Bark-Skala einen linearen Zusammenhang zur Länge der Basilarmembran aufweisen und die Frequenz ab ca. 500 Hz eine grob logarithmische Unterteilung (siehe [FZ07]).

Einschwingzeit Die *Einschwingzeit* stellt den *wahrnehmbaren Anfangsmoment des Einschwingvorganges* (engl. *onset*) eines Klages dar, d.h. der Beginn eines akustischen Ereignisses, z.B. eine gespielte Musiknote oder ein Paukenschlag (vgl. [Kla99]). In den meisten Fällen ist die Einschwingzeit mit dem Anfang des *Einschwingvorganges* (engl. *transient*) gleich [BDA⁺05], welcher den am deutlichsten wahrnehmbaren Bereich des Audiosignales darstellt. Er enthält die An klingphase (engl. *attack*) und den Hauptteil der Ab klingphase (engl. *decay*).

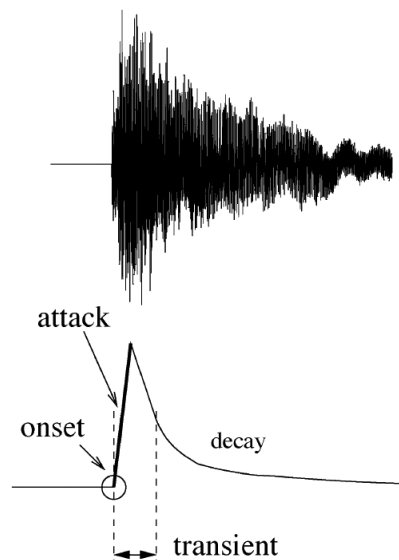


Abbildung 3.10: Relation der Konzepte *wahrnehmbarer Anfangsmoment des Einschwingvorganges*, *Einschwingvorgang*, *An klingphase*, *Ab klingphase*

In Abbildung 3.10 werden die Zusammenhänge der einzelnen Merkmale an Hand eines Audiosignals einer einzelnen Note grafisch verdeutlicht. Die Einschwingzeit bildet somit die Grundlage für weitere höherwertige Eigenschaften, z.B. *Tatum*, Grunds Schlag oder Rhythmus (siehe Abschnitt 3.1.2.2 Paragraph „Rhythmus“), und segmentiert somit ein komplexes Signal, wie es in der Musik vorkommen kann.

Lautheit Die *Lautheit* (N) ist eine psychoakustische Größe zur Abbildung des menschlichen Lautstärkeempfindens. Sie gibt die subjektive Beurteilung der Intensität des Klages wieder [Jeh05]. Als Größeneinheit wird Sone verwendet, wobei 1 sone die empfundene Lautstärke eines Schallereignisses bei einem Lautstärkepegel von 40 phon ist. Eine Verdopplung der Lautheit entspricht einer als doppelt so laut empfundenen Lautstärke (siehe [Lid06] und [dlMHR05]).

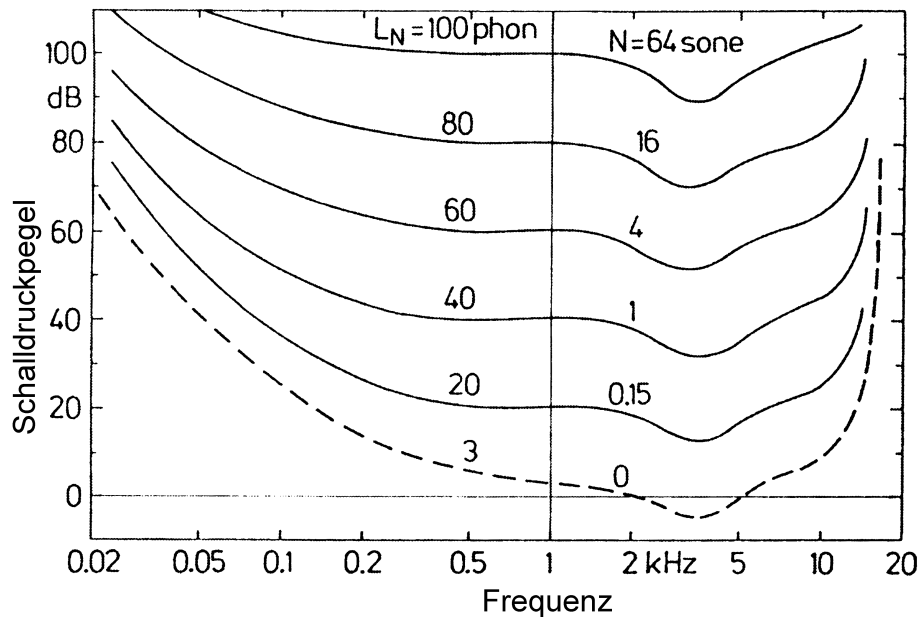


Abbildung 3.11: Referenzkurven von gleichen Lautstärkepegeln (ungefähre Lautheitsumrechnung) [FZ07]

Dem gegenüber steht die Lautstärke als objektive physikalische Maßeinheit der Amplitude des Schalls mit Schalldruckpegel in Dezibel (dB) (siehe Abschnitt 3.1.1.1 Paragraph „Intensität“) und dem zugehörigen psychoakustischen Vergleichsmaß Lautstärkepegel in Phon. Schalldruckpegel und Lautstärkepegel stimmen bei einem Sinuston mit einer Frequenz von 1000 Hz überein. Abbildung 3.11 zeigt zum Vergleich Referenzkurven der Wahrnehmung von gleichen Lautstärkepegeln (z.B. 40 phon) von reinen Tönen bei unterschiedlichen Frequenzen und ihren zugehörigen Lautheitswerten.

Die Wahrnehmungsschwellen variieren stark in Bezug auf die Frequenz. Niedrig- und hochfrequente Schallereignisse werden erst bei höheren Schallintensitätspegeln wahrgenommen. Die gebräuchlichen Frequenzen von Musik und Sprache befinden sich deutlich innerhalb des Bereiches von Hörschwelle bis Schmerzgrenze und sind in Abbildung 3.12 durch jeweils gestreifte Abschnitte dargestellt.

Es existieren verschiedene Modelle zur Beschreibung der Lautstärkewahrnehmung des menschlichen Gehörs, z.B. von Zwicker (DIN 45631) oder Moore und Glasberg (siehe [Eus09] und [Jeh05]). Über die jeweiligen Modelle kann der Lautstärkepegel aus der Lautheit berechnet werden (siehe [Hau09]).

Generell hängt die Lautheitsempfindung aber von der Tondauer der Klänge bzw. von den Frequenzen der überlagernden Töne ab (siehe [Roe00]). Bei Tönen gleicher Frequenz oder Frequenzgruppe addieren sich die einzelnen Intensitäten (siehe Abschnitt 3.1.1.1 Paragraph „Intensität“) einfach und die Lautheit resultiert dann aus der Gesamtintensität. Wenn die Frequenzen der überlagernden Töne nun die Breite der Frequenzgruppe überschreiten, müssen noch die Maskierungseffekte der einzelnen Töne mit einberechnet werden (siehe [Roe00]).

Schwankungsstärke und Grobheit *Schwankungsstärke* (engl. *Fluctuation Length* in *vacil*) und *Grobheit* (auch *Rauhigkeit* genannt; in *asper*) sind zwei auditive Sinnes-

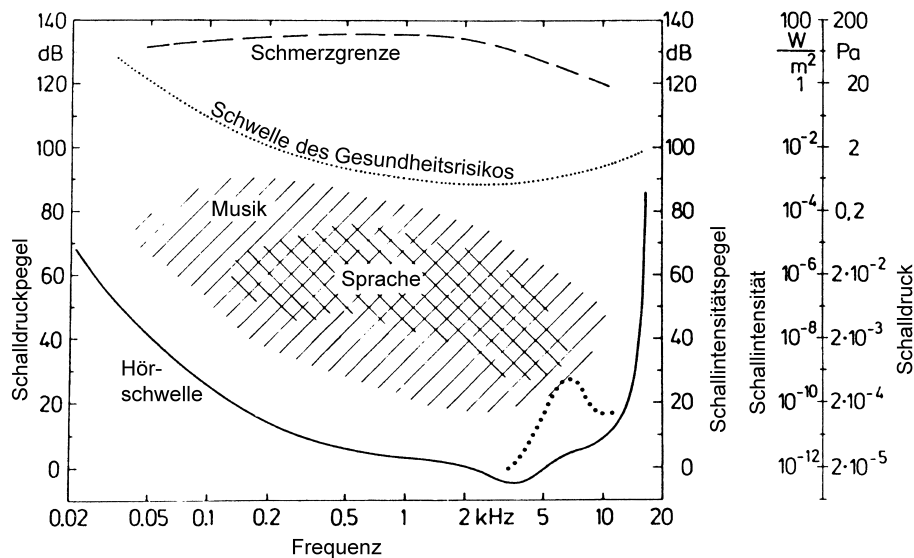


Abbildung 3.12: Durchschnittlicher menschlicher Hörbereich mit den Abschnitten für Musik und Sprache [FZ07]

empfindungen, welche sich auf die Amplitudenschwankungen der Lautheit¹¹ (auch *Modulationsfrequenzen* genannt; siehe Abbildung 3.13) beziehen. Die Schwankungsstärke ist dabei grob bis zu einer Modulationsfrequenz von 15 – 20 Hz wahrnehmbar und die Grobheit darüber (siehe [FZ07]).

Ihr Maximum hat die Schwankungsstärke bei einer Amplitudenschwankung von ca. 4 Hz, wovon sich auch das Maß *Vacil* bzw. die Festlegung 1 *vacil* ableitet (ein 60-dB, 1-kHz Ton mit einer 100%-igen 4-Hz-Modulationsfrequenz). Da der Modulationsfrequenzbereich der Schwankungsrate auch vollständig den üblichen Tempobereich der Musik (siehe Abschnitt 3.1.2.2 Paragraph „Rhythmus“) mit einschließt, eignet sich diese Größe auch gut für Berechnungen der Rhythmus- und Grundschlagerkennung und dort im Speziellen zur Einbeziehung der *zeitlichen Maskierung* (siehe Abschnitt 5.1.5.1 und Abschnitt 5.1.3 Paragraph „Maskierungsauswertung“). Dafür entwickelte Zwickler et al. [FZ07] eine auf Frequenzgruppen basierende Bestimmung der Schwankungsrate.

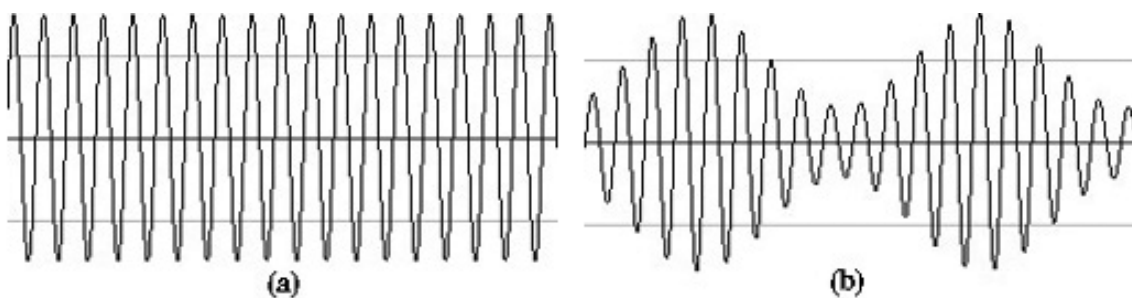


Abbildung 3.13: Klangsignal mit (a) fester Amplitude und (b) Amplitudenschwankungen [Vas05]

Die *Grobheit* wurde 1863 von Helmholtz eingeführt. Sie beschreibt grelle, kratzige und raue Klänge (z.B. eine von 2 Flöten gespielte *kleine Sekunde*), welche bei Signalen mit

¹¹d.h. die zeitlichen Schwankungen des Schalls

Amplitudenschwankungen von ca. 20 bis 200 Hz vorkommen (je nach Tonhöhenbereich) [Vas05]. Die Rauigkeit von 1 *asper* entspricht einem 60-dB, 1-KHz Ton, welcher zu 100% in einer Amplitude bei einer Modulationsfrequenz von 70 Hz moduliert ist [FZ07]. Folgende Attribute nehmen Einfluss auf die Grobheit: Amplitudenschwankungsrate, Grad der Amplitudenschwankung¹² und Trägerfrequenz.

Grobheit wird auch als *sensorische Dissonanz* bezeichnet (siehe [BVL⁺04]), wobei die empfundene Dissonanz sich auf die Ästhetik der Musiktheorie der *westlichen Musik* bezieht (siehe Abschnitt 3.1.2.2 Paragraph „Harmonie“). In anderen Bereichen wird das Mittel der Amplitudenschwankungen bewusst eingesetzt um mit der Grobheit kreativ zu arbeiten, z.B. um Klangfarbenvariationen eines Instrumentes zu erzeugen (siehe [Vas05]). Deshalb gilt die Grobheit auch als ein wesentliches Attribut der Klangfarbe (siehe Abschnitt 3.1.2.1 Paragraph „Klangfarbe“).

Die Gründe für die Wahrnehmung der Grobheit liegen im Wesentlichen in den mechanischen Eigenschaften des Innenohres, welches „versucht“ den ankommenden Schall in Nervenimpulse umzusetzen. Durch eine relativ hohe Amplitudenschwankungsrate kann der Hörfrequenzanalysemechanismus das Eingangssignal nicht richtig auflösen, was zu einer Instabilität der Basilarmembran (als Teil der Hörschnecke des Innenohrs) führt (vgl. [Vas05]). Die Sinnesempfindung der Grobheit hängt stark mit den schon definierten Frequenzgruppen zusammen (siehe Einführung dieses Abschnittes). So wird z.B. ein komplexer Hörreiz von zwei reinen Tönen, deren Frequenzen sich innerhalb einer Frequenzgruppe befinden, als Rauigkeit empfunden und das Gegenteil als *Glattheit* (siehe [Roe00]).

Es existieren zahlreiche Modelle zur Schätzung von Grobheit. Vassilaki hat im Rahmen seiner Doktorarbeit ein Bewertungsmodell für Grobheit entwickelt, was auch komplexe Spektren behandeln kann (siehe [Vas01]).

3.1.2 Angereicherte Audiosignaleigenschaften

Angereicherte Audiosignaleigenschaften werden von physikalischen (Abschnitt 3.1.1.1) und sinnesempfundene (Abschnitt 3.1.1.2) Merkmalen abgeleitet und auch als *globale* bzw. *kontextuelle Deskriptoren* bezeichnet (siehe [Les06]), welche die *musikalischen Wahrnehmung* bilden (vgl. [dIMHR05]). Sie beziehen sich auf die interne subjektive Musikerfahrung, d.h. die Charakteristiken können *werte-beladen* (z.B. Emotionen oder Musikgenre, siehe Abschnitt 3.2) oder *nicht-werte-beladen* (z.B. Tonhöhe oder Klangfarbe, siehe folgender Abschnitt 3.1.2.1) sein. *Werte-beladene* Aussagen beziehen sich hierbei auf die Meinung, Überzeugung oder das Gefühl Einzelner [Les06].

Im Folgenden wird nun auf die perzeptuellen Eigenschaften einzelner Klänge eingegangen und das Zusammenwirken dieser danach durch die raum-zeitlichen Struktureigenschaften beschrieben. Generell können die Werte diese Eigenschaften eine doppeldeutige oder unklare Bedeutung haben, welche durch die Wahrnehmung eines jeden einzelnen Hörers und dessen Umgebung beeinflusst wird.

3.1.2.1 Perzeptuelle Eigenschaften

Die durch die Hörsinnesempfindungen vom menschlichen Hörsystem (siehe Abschnitt 3.1.1.2) erfassten Klänge werden anschließend vom Gehirn weiterverarbeitet und inter-

¹²Abgeleitet von der Differenz der Höhen und Tiefen eines Signals

pretiert (siehe Abbildung 3.4). Je nach Vorwissen im Bereich der Musikwahrnehmung und -interpretation kann ein Hörer die folgenden perzeptuellen Eigenschaften in einen Kontext einordnen und unterscheiden.

Tonhöhe Die *Tonhöhe* (engl. *pitch*) ist die *wahrgenommene Qualität eines Klanges*, welcher hauptsächlich eine Funktion seiner *Grundfrequenz* ist [Dow03]. Die bekannteste Darstellungsform ist die vertikale Position einer Musiknote in Bezug zu Notenlinien. Verschiedene Repräsentationsformen haben sich entwickelt, u.a. Notennamen (z.B. C, E, Cis) oder *Tonhöhenklassen* (0 bis 11).

Der Höhenunterschied (*Frequenzverhältnis*) zwischen zwei Tönen wird *Intervall* genannt und in *Halbtönen* angegeben. 12 aufeinander folgende Halbtöne, d.h. 12 verschiedene Tonhöhen, ergeben dabei eine *reine Oktave*, wie sie innerhalb der westlichen Musiktheorie definiert wird (auch bekannt als *chromatische Skala*; vgl. [Hai00]). Die Oktave ist nach dem *Oktavraum* der *diatonischen Tonleiter*¹³ benannt, welcher 8 Tonstufen bzw. 7 Stammtöne umfasst (siehe [mis09c]). Eine Verdopplung der Frequenzen der enthaltenen Tonhöhen führt zu einem neuen Oktavraum. Musikinstrumente und Stimmlagen haben dabei oft eine *Tonhöhenreichweite*, welche sich über mehrere Oktavräume hinweg erstreckt.

Abbildung 3.14 zeigt eine Auswahl verschiedener Musikinstrumente und Stimmlagen und die dazugehörigen Frequenzumfänge, wobei manche Tonhöhenreichweiten gar nicht vollständig erfasst sind, z.B. die der Orgel, mit der auch noch tiefere Töne möglich sind. Weitere spezielle Intervalle, d.h. mit besonderen Frequenzverhältnissen, heißen z.B. *Sekunde* oder *Terz*, wobei es davon wiederum verschiedene Abwandlungen existieren, z.B. *große Sekunde* oder *kleine Terz*.

Ein weitere Begrifflichkeit, die eng verwandt mit der Tonhöhe ist, ist die *Tonart*, welche nach dem *Grundton* (dem ersten Ton einer Tonleiter) und dem, in der westlichen Musiktheorie gebräuchlichen, *Modus*¹⁴ (auch Kirchtöneleiter [Hem01]) benannt wird. Eine Teilmenge der Kirchtöneleitern sind die beiden häufig verwendeten *Tongeschlechter Dur* (*ionischer Modus*) und *Moll* (*äolischer Modus*). Diese Vereinfachung wird auch in der gegenwärtig vorherrschenden *Dur-Moll-Tonalität* wiedergegeben (siehe Abschnitt 3.1.2.2 Paragraph „Harmonie“ und [mis09c]). In Abbildung 3.15 ist eine Tonleiter mit der Tonart *C-Dur* illustriert.

Der wiederkehrende Zyklus von ähnlich klingenden Tönen bekommt durch den Rahmen der Oktaven den Begriff der *Oktavenähnlichkeit*. Auf diese Ähnlichkeit, die in allen Kulturen der Welt vorrangig anerkannt ist, begründet sich auch die *Zweikomponententheorie* der Tonhöhe (siehe [dlMHR05]). Wobei sich die eine Komponente auf den wiederkehrenden zyklischen Toncharakter bezieht und *Tonigkeit*, *Chroma* aber auch (*Ton-*)*Qualität* genannt wird. Die andere Komponente gibt die Höhe im Tonraum an und bezieht sich dabei auf die linear ansteigende Höhe eines Tones. Sie wird (*Ton-*)*Helligkeit* oder auch direkt nur *Tonhöhe* genannt. Der Zusammenhang dieser beiden Komponenten lässt sich recht gut in einer Schraubenlinie¹⁵ darstellen, deren Zyklus man auf einen Kreis projizieren kann (siehe Abb. 3.16).

¹³Tonleiter als eine Stufenfolge bzw. Auswahl von Tönen

¹⁴Daneben gibt es noch andere Tonleistersysteme

¹⁵u.a. von Moritz Wilhelm Drobisch 1852 vorgeschlagen

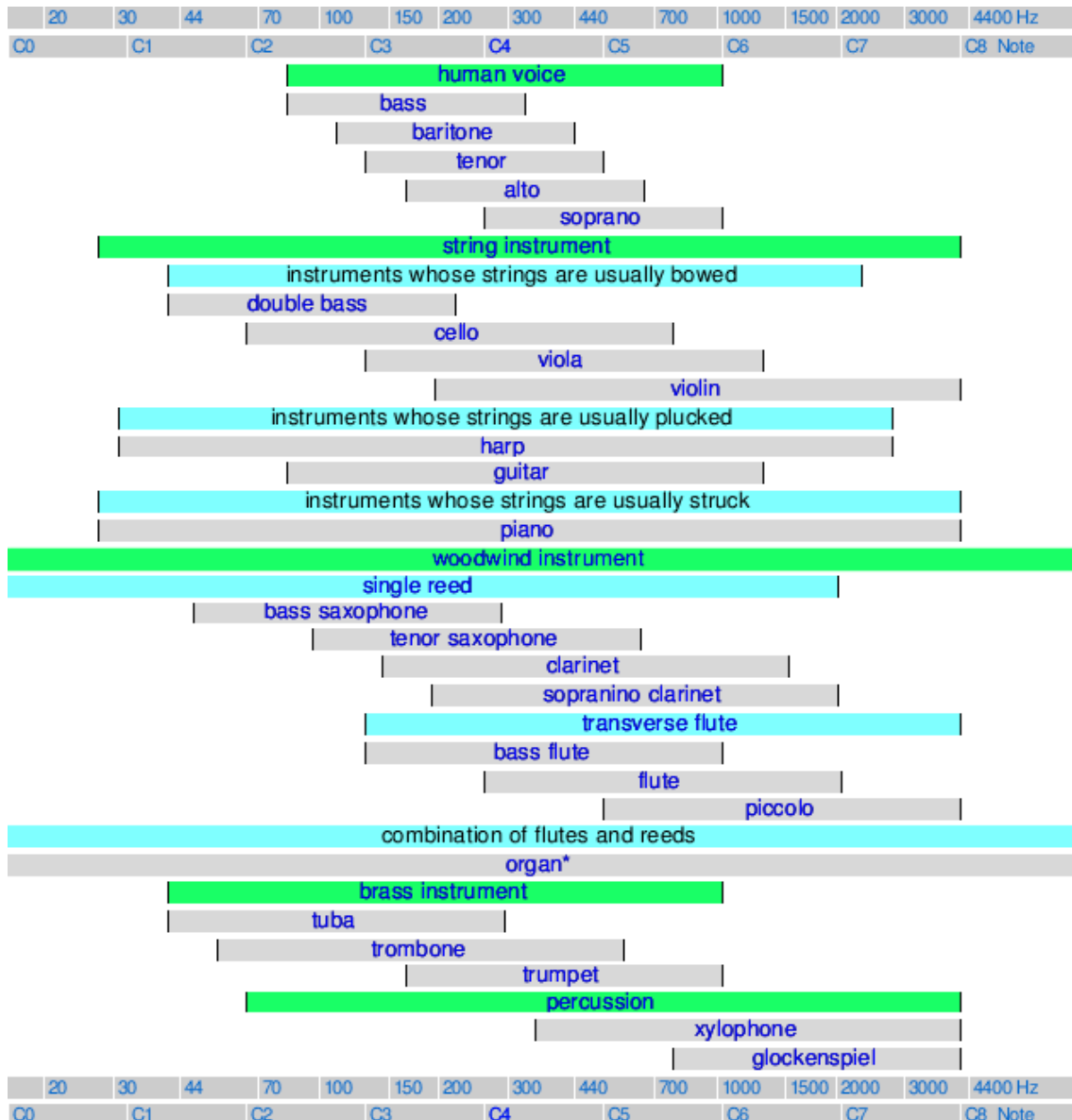


Abbildung 3.14: Tonhöhenreichweite verschiedene Musikinstrumente und Stimmlagen in Bezug auf eingeschlossene Oktavräume und Frequenzumfang (Auszug von [mis09h])

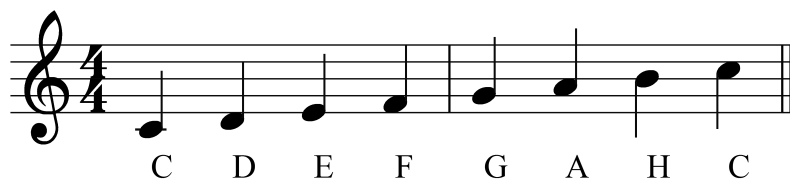


Abbildung 3.15: C-Dur-Tonleiter in einem Notensystem

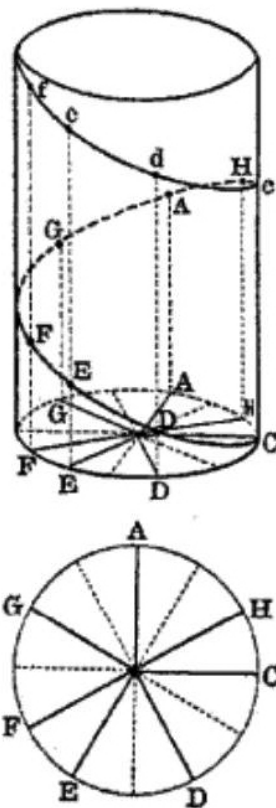


Abbildung 3.16: Schraubenlinienrepräsentation der Helligkeit und Tonqualität der 12 chromatischen Töne [HW30]

Tondauer Die *Tondauer* ist ein musikalischer Parameter, welcher die wahrgenommene zeitliche Länge eines Tones oder Klangs angibt. In diskreter Form angegebene *Notenwerte* (siehe Abb. 3.17), wie z.B. eine *Viertel-* oder *Achtelnote* (als gespielter Ton) bzw. *Ganze* oder *Halbe Pause* (als Ruhemoment), werden relativ zu einem bestimmten Tempo (siehe Abschnitt 3.1.2.2 „Paragraph“) wiedergegeben und je nach Lautstärke und Hörempfinden als unterschiedlich lang vom auditorischen System erfasst.

Name	Note	Pause	Name	Note	Pause
Ganze			Sechzehntel		
Halbe			Zweiunddreißigstel		
Viertel			Vierundsechzigstel		
Achtel					

Abbildung 3.17: Noten- und Pausewerte in der gebräuchlichen Notation der westlichen Musiktheorie [Hem01]

Diese Eigenschaft kann mit Einheiten der Zeit (siehe Abschnitt 3.1.1.1 Paragraph „Zeit und Dauer“) absolut und exakt gemessen werden. Dabei liegt die minimal empfindbare Tondauer bei ca. 10–15 ms, und ab ca. 15 ms kann auch zwischen verschiedenen

Tonhöhen (siehe vorheriger Paragraph) und Lautstärken (siehe Abschnitt 3.1.1.2 Paragraph „Lautheit“) unterschieden werden (siehe [Roe00]). Wichtig bei der Wahrnehmung der Tondauer eines Tones oder Klangs sind auch Klangereignisse, welche davor, gleichzeitig oder danach eintreten und somit Einfluss auf die Separierung haben.

Klangfarbe *Klangfarbe* ist die Eigenschaft bzw. Sinnesempfindung, die es ermöglicht zwischen Klängen von ein und derselben Lautheit und Tonhöhe, erzeugt von Musikinstrumenten oder menschlichen Stimmen, zu unterscheiden, d.h. sie bezieht sich auf den charakteristischen Klang eines oder mehrerer Musikinstrumente bzw. Stimmenlagen (z.B. Tenor; vgl. [Jeh05], [dlMHR05] und [Les06]). Beschrieben wird dieses Merkmal durch sein Frequenzspektrum (siehe Abschnitt 3.1.1.1 Paragraph „Spektrum“) und dessen Veränderung im zeitlichen Verlauf (Spektralfluss).

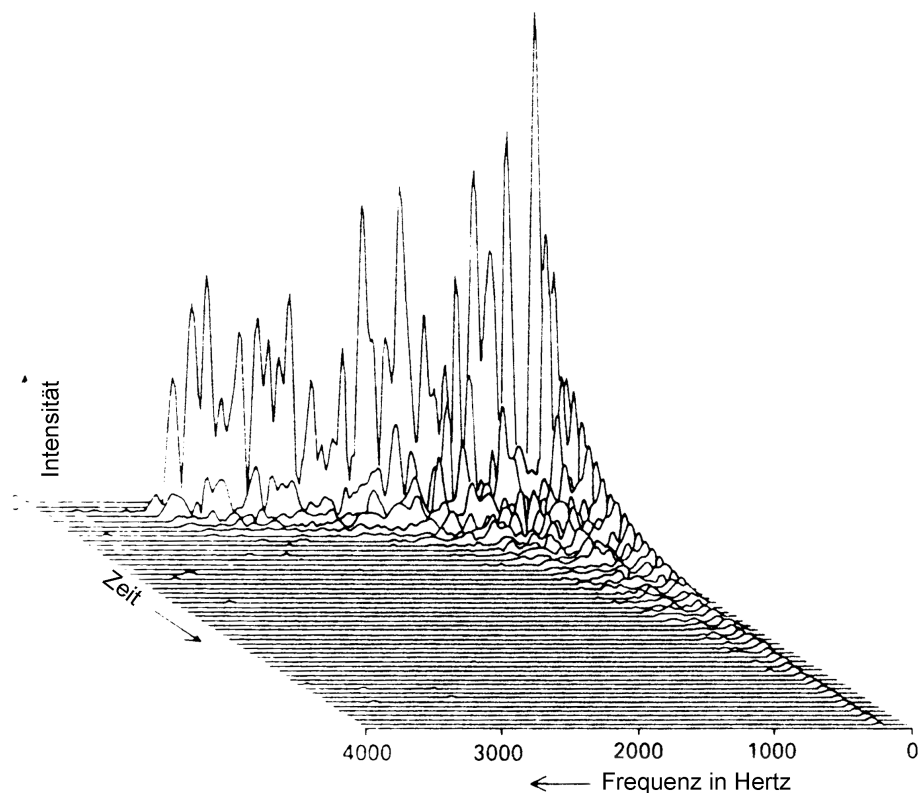


Abbildung 3.18: Spektrogramm eines Tom-Tom-Schlages [Hem01]

Wo bei der Tonhöhe der Grundton, d.h. die Grundfrequenz, von wesentlicher Bedeutung ist, bestimmen die *Partialtöne*, im Speziellen die *Obertöne* (auch *Harmonische* genannt) bzw. *Untertöne* (als ganzzahlige Vielfache bzw. Teiler der Grundfrequenz), die Lautheit bzw. Grobheit (siehe Abschnitt 3.1.1.2 Paragraph „Lautheit“ bzw. „Schwankungsstärke und Grobheit“) und auch Rauschen die Klangfarbe. Darüber hinaus ist die Anlingphase (siehe Abschnitt 3.1.1.2 Paragraph „Einschwingzeit“) ein entscheidungstragender Zeitabschnitt zur Charakterisierung (siehe [Jeh05]). Viele Musikinstrumente, vor allem Streichinstrumente und Blasinstrumente, weisen eine Obertoncharakteristika auf. Untertöne kommen in metallischen Klängen vor, wie sie z.B. von einer Glocke erzeugt werden. Schlagzeugklänge enthalten zusätzlich noch Rauschanteile. In Abbildung 3.18 ist ein Frequenzspektrum eines Tom-Tom-Schlages aufgezeichnet, welches in der Anschlagphase noch recht komplex ist, sich aber dann auf eine einfache Schwingung reduziert.

Neben der eigentlichen Form der verschiedenen Musikinstrumente hat die Spielweise (z.B. *fortissimo*) Einfluss auf die Wahrnehmung der Klangfarbe (siehe [Dow03]). Durch z.B. sanftes oder hartes Anspielen ergibt sich eine wesentlich unterschiedlichere Hüllkurve des Klanges, mit anderen Anklingphasen, Lautheiten und Tondauern.

3.1.2.2 Raum-zeitliche Struktureigenschaften

Raum-zeitliche Struktureigenschaften beschreiben das Zusammenwirken von Klängen auf verschiedene Weise. Mit Harmonien wird versucht eine (subjektive) Schönheitsbewertung von gleichzeitig und nacheinander erklingenden Tönen zu definieren. Der Rhythmus hingegen gibt ein zeitliches Grundkonzept für darauf gelegte Abfolgen von ein oder mehreren Tonhöhen, der Melodie. Letztendlich können dabei auch noch Variationen, der festgelegten Muster der letzten beiden genannten Eigenschaften erzeugt werden, welche unter *Musikvortragsbezeichnungen* definiert werden.

Harmonie Die *Harmonie*, als eine bedeutende charakteristische Eigenschaft der westlichen Musik, beschreibt das *Zusammenpassen*¹⁶ von elementaren Klangformen und klanglichen Konstellationen [Sau07]. Folgende Merkmale werden oft als Grundlage zur Beschreibung von Harmonie genutzt: *Akkord*, *Tonalität*, *Progression*, *Modulation* und *Konsonanz* oder *Dissonanz*.

Akkorde lassen sich i. A. als gleichzeitiges Erklingen von zwei oder mehreren Klängen unterschiedlicher Tonhöhe definieren, und üblicherweise durch ihre Intervalle und *Grundtonart* beschreiben (vgl. [mis09c]). Eine genauere Festlegung setzt aber mindestens drei gleichzeitig erklingende Tonhöhen voraus mit einem Intervall von 3, genannt *Terz* (siehe [Les06]). Die wohl häufigste Form eines Akkordes ist der *Dreiklang*, welcher sich aus den Grundton und zwei Terzen zusammensetzt. Abgebildet auf die Tonleiter, welche auf Basis des Grundtones definiert wird, entspricht dies dem ersten, dritten und fünften Ton dieser Tonfolgendefinition (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“).

Tonalität befasst sich im weitesten Sinne mit der systematischen Anordnung von Tönen und deren Beziehung untereinander und ist damit ein Organisationssystem zur Harmoniebeschreibung (vgl. [mis09c] und [Les06]). Sie wird heutzutage als *Harmonie der Tonart* aufgefasst, obwohl man sich bei dieser Definition typischerweise nur auf das Harmonisieren von *Dur-* und *Moll-Dreiklänge* bezieht¹⁷. Diese Dreiklänge stehen in einem *Quintverhältnis*, d.h. ein Frequenzverhältnis von 3:2 zwischen dem höchsten und dem tiefsten Ton (der Grundton) des Dreiklages (siehe [Sau07]). Der Unterschied zwischen einem *Dur-* und einem *Moll-Dreiklang* besteht in dem Frequenzverhältnis der auf den Grundton folgenden *Terze*. Wo bei einem *Dur-Dreiklang* das auf den Grundton folgende Intervall erst eine *große Terz*, mit dem Verhältnis 5:4, und dann eine *kleine Terz*, mit dem Verhältnis 6:5, ist, ist diese Anordnung bei einem *Moll-Dreiklang* genau umgekehrt (erst die *kleine Terz* und dann die *große Terz*; siehe [Sau07]).

Die grundlegenden Bestandteile der Harmoniebeschreibung von *Tonalität* sind die *konsonanten Dreiklänge Tonika*, *Dominante* und *Subdominante*, welche sowohl aufeinander folgend (*Progression*) als auch zusammen gespielt werden können. Aus dem Grundton

¹⁶Als eine Beschreibung von wahrgenommenen Eigenschaften von Objektbeziehungen unter dem Prinzip der Ästhetik (siehe [Sau07])

¹⁷Deshalb wird Musik auch als *tonal* bezeichnet, wenn sie nur auf *Dur* und *Moll* basiert und ihr Gegenteil mit *atonal*

der *Tonika* bilden sich für dieses System die Tonart der Tonleiter, von welcher der vierte bzw. der fünfte Ton jeweils der Grundton für *Subdominante* bzw. *Dominante* ist (siehe [mis09c]). Somit werden die sieben *Stammtöne* dieser Tonart komplett abgedeckt.

2:1	<i>Oktave</i>
3:2	<i>Quinte</i>
5:4	<i>große Terz</i>
6:5	<i>kleine Terz</i>

Tabelle 3.1: Auf *Konsonanz* hindeutende Frequenzverhältnisse (siehe [Sau07])

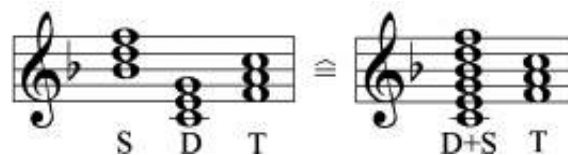


Abbildung 3.19: *Tonika* (T), *Dominante* (D) und *Subdominante* (S) von *F-Dur* in *Konsonanz* und *Dissonanz* [Sau07]

Die Progression, als eine nicht näher beurteilte Abfolge von Akkorden, wird durch die (subjektiv festgelegte) Schönheitsdefinition *Konsonanz* bewertet, welche die Akkorde in ihrer Form als Wohlklang einordnen kann. Ausschlaggebend für diese Einschätzungen der „in sich ruhenden Klänge“ [Les06] sind die Proportionen der Intervalle der zusammenklingenden Töne (siehe Tabelle 3.1). Die zeitliche Anordnung von *Tonika*, *Dominante* und *Subdominante* wird daher speziell als *Kandenz* bezeichnet (siehe [Sau07]). Wenn man diese Dreiklänge oder Teile davon übereinander lagert, können Dissonanzen entstehen, welche auch als *Grobheit* wahrgenommen werden (siehe Abschnitt 3.1.1.2 Paragraph „Schwankungsrate und Grobheit“). Dissonanz bildet somit das wahrgenommene Gegenteil zu Konsonanz. In Abbildung 3.19 werden die drei Akkorde von *F-Dur* bzw. *d-Moll* (jede Durtonart besitzt eine parallele Molltonart) in konsonanter bzw. dissonanter Form gegenübergestellt. Eine *Modulation* ist letztendlich der harmonische Übergang von einer Tonart zu anderen [Sau07]. Dies geschieht im einfachsten Fall mit Hilfe der Paralleltonart von *Tonika*, *Dominante* oder *Subdominante* (siehe [mis09c]).

Der *tonalen Harmonik* steht die *modale Harmonik* gegenüber, welche auch im Jazz (z.B. Stücke von Miles Davis) und in der Popmusik (z.B. Stücke von *The Beatles*) genutzt wird. Sie beruht auf den Kirchtonleitern und gilt als freie Folge (Progression) von Akkorden des gewählten Modus (siehe [Hem01]).

Rhythmus Ein Rhythmus kann als Arrangement von *Dauermustern* (engl. *Durational Pattern*, siehe [DWW⁺75]) oder variierenden Tondauer definiert werden [Les06]. Er gibt damit der Musik eine Zeitstruktur, welche sich von den Merkmalen *Grunds Schlag* (engl. *beat*), *Tempo*, *Takt*, und *Metrum* charakterisieren lässt. Abbildung 3.20 zeigt die Zusammenhänge der einzelnen Eigenschaften (ohne Tempo).

Der Grunds Schlag (auch Pulsschlag, siehe [Wie17]) wird als herbeigeführter periodischer Impuls wahrgenommen und beruht auf einer kognitiv komplexen Sinnesempfindung [Jeh05]. Er ist eine heruntergerechnete (engl. *down-sampled*), angeordnete Version des

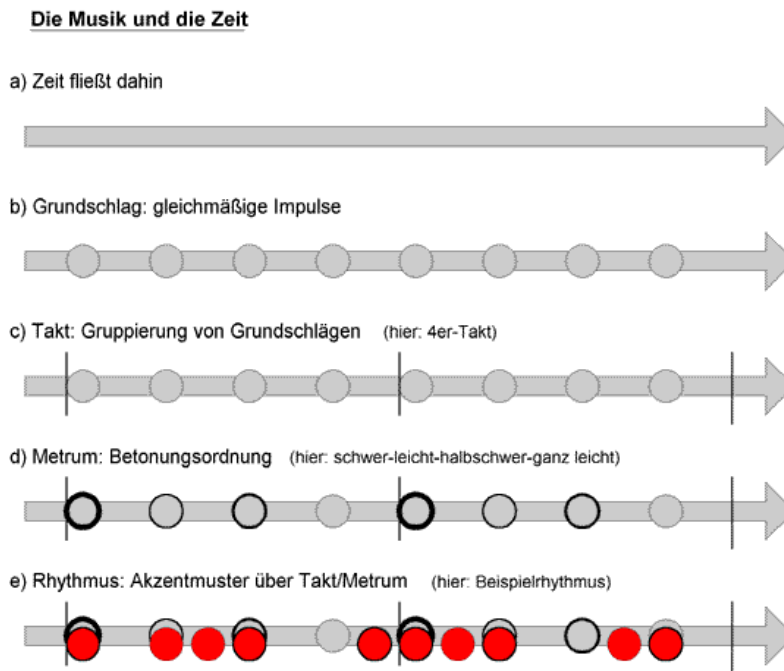


Abbildung 3.20: Gegenüberstellung der verschiedenen zeitstrukturierenden Merkmale der Musik [Qpa09]

Tatum, d.h. ein ganzzahliges Vielfaches dieses (siehe [Bil93]). Ein *Tatum* ist die niedrigste geordnete Impulsfolge, welche ein Zuhörer intuitiv aus den Zeitintervallen der wahrgenommen Musikereignisse (z.B. der gespielten Töne) schließen kann, z.B. die kürzeste Note [Jeh05]. Es ist grob mit der Zeiteinteilung der Einschwingzeiten gleich zu setzen (siehe 3.1.1.2) und definiert die Qualität dieser gewonnenen Audiosignaleigenschaft, wobei hier auch schneller werdende (*Accelerandos*) und langsamer werdende (*Ritardandos*) Grundschatläge erfasst werden.

Eng in Verbindung zu der Basiszeiteinheit der Musik, dem Grundschatlag, steht das Tempo, welches eine Geschwindigkeitsreferenz auf diese darstellt. Üblicherweise wird diese Maß in *Grundschatläge pro Minute* (engl. *beats per minute*; BPM) angegeben und variiert vor allen in dem Bereich zwischen 40 und 260 BPM, wobei das Mittel bei ca. 120 BPM¹⁸ liegt (siehe [Jeh05]). Subjektive Tempoempfindungen werden durch Tempozeichnungen wie z.B. *adagio* (langsam), *moderato* (mäßig bewegt) oder *presto* (sehr schnell) angegeben (siehe [Hem01]).

Unabhängig vom Tempo ist die Strukturierungseinheit für Grundschatläge, der Takt, welcher diese zu gleichen Gruppierungen von *Elementen* durch *Taktstriche* unterteilt (siehe [Wie17][Hai00]). Wo beim Tempo die Zeitangaben als absolut gelten, sind sie beim Takt als relativ anzusehen. Somit kann ein vorgegebener Takt (theoretisch) in jedem Tempo gespielt werden. Die in der Popmusik wohl am häufigsten genutzte *Taktart* ist der *4/4-Takt* (vgl. [GCD07]). In der einfachsten Variante besteht ein solcher Takt aus vier Viertelnoten, d.h. allgemein müssen die Tondauern der genutzten Töne (oder Pausen) insgesamt immer wieder die des Taktes ergeben. Eine einfache Einteilung der Taktarten ist die Unterscheidung zwischen *geraden Taktarten*, z.B. *4/4-* oder *6/8-Takt* und *ungeraden Taktarten*, z.B. der bekannte *3/4-Takt* als Grundlage des Walzers. Dabei haben die *Grundtaktarten*

¹⁸was einer 2 Hz-Modulationsfrequenz entspricht

Zähler von 2 (*zweizeitig*) oder 3 (*dreizeitig*), und die *zusammengesetzten Taktarten* sind entweder Vielfache oder Kombinationen (z.B. *5/4-Takt*) dieser beiden Grundformen (siehe [Wie17]).

Unter einem *Metrum* (vom griech. *metron*) versteht man in der Musik eine Organisationsform für *Betonungsmuster* von Grundsschlägen (Tatums), welche dadurch die Hintergrundstruktur eines Rhythmus vorgeben. Oft basiert dieses musikalische Ordnungsprinzip auf der Musikstruktureinheit Takt. Betonungen werden häufig in 5 Charakteristiken unterteilt werden¹⁹: *unbetont*, *ganz leicht*, *halbschwer* und *schwer* und auf die jeweiligen *Segmente* (ganze Takte bzw. *Zählzeiten* davon) bzw. Elemente des Metrums angewandt werden (siehe Abb. 3.20). Die vom Takt her bekannten Unterteilungen wurden von den jeweiligen Ordnungen (*gerade* und *ungerade Ordnung*) der Metrik abgeleitet, wie auch die dazugehörigen *einfachen* und *kombinierten Formen* (siehe Grundtaktart und zusammengesetzte Taktart). Wenn bei den kombinierten Formen gerade und ungerade Ordnung aneinander gefügt werden, kann dies auf eine vielfältige Weise passieren. Somit entstehen hierbei die *unregelmäßigen Metren* (siehe [Wie17]). Ein *fünfzeitiges Metrum*, kann sich aus den Kombinationen 3+2 und 2+3 ergeben.

Auf Basis der gegebenen Hintergrundstruktur durch Metrum und Takt, können Töne verschiedener Tondauern gesetzt und variiert werden, welche dadurch die sogenannten *Dauermuster* ergeben. Diese Dauerfolgen können so komplex aufgebaut sein (z.B. *Synkopen*²⁰), dass es für ungeübte Zuhörer schwer fällt, den eigentlichen Grundschatz aus dem Rhythmus herauszuhören. Doch gerade diese Variationen von Tondauer und Betonung in Bezug auf den Grundschatz erzeugen die Spannung des Rhythmus (vgl. [Hem01] und [Hai00]).

Melodie Eine *Melodie* kann als eine Menge von Tonhöhen oder Intervallen verstanden werden, welche zeitlich geordnet wahrgenommen werden [Dow03]. Ein Verlauf einer Melodie wird also durch die jeweiligen Tonhöhen der gespielten Töne, welche durchaus auch gleichzeitig als Akkord gespielt werden können, und deren Tondauer charakterisiert. Somit können sie mit Mitteln der Harmonie und des Rhythmus konstruiert und analysiert werden. Insgesamt soll eine solcher Tonverlauf eine wiedererkennbare Einheit ergeben, welcher aber auch variiert werden kann, z.B. im Bereich der Klassik und des Jazz (siehe [Les06]).

Von wahrnehmbarer Bedeutung sind die beiden Merkmale *Melodiekontur* und *Melodiebereich*. Auf der einen Seite beinhaltet die Melodiekontur (auch Melodiekurve) den Tonhöhenverlauf der gespielte Töne, welcher steigen, fallen oder gleich bleiben und z.B. durch den *Parsons Code* (siehe [Par75]) beschrieben werden kann. Auf der andere Seite gibt der Melodiebereich das Intervall der höchsten und tiefsten Tonhöhe der Melodie an. Dabei haben Melodiebereiche mit einer großen Reichweite eher einen animierenden Charakter und wirken sich auch auf die Entwicklung der Melodiekontur aus (siehe [Les06]).

Musikvortragsbezeichnungen Abwechslung, d.h. Hervorhebung und Dämpfung, in Melodie und Rhythmus bringen die *Akzente* der *Musikvortragsbezeichnungen* (auch *Musikvortragsveränderungen*). Sie unterteilen sich in die *Tempo-* (*Agogik*) und *Lautheitsänderungen* (*Dynamik*). Beide Kategorien können oder sollen Auswirkungen auf die

¹⁹Betonungstaffelungen mit einer anderen Anzahl von Ebenen sind auch möglich

²⁰Verschiebung der Betonung auf eine vorher unbetonte Zählzeit

bis zur Anwendung der *Musikvortragsbezeichnung* vorherrschende emotionale Stimmung erzeugen und beziehen sich somit relativ auf die gegebenen Eigenschaften des wahrgenommenen Audiosignals (vgl. [Les06]). Die Abwechslungen können sofort oder verzögert auftreten und wahrgenommen werden.

Auf der einen Seite werden die plötzlichen *agogischen Akzente* durch Terme, wie z.B. *sofort langsamer sein* (ital. *ritenuto*), angegeben. Auf der anderen Seite werden die allmählichen Tempoänderungen mit Begriffen, wie z.B. *schneller werdend* (ital. *accelerando*) oder *langsamer werdend* (ital. *ritardando*), beschrieben.

Analog gibt es für plötzliche *dynamische Akzente* Beschreibungen, welche mit den eigentlichen *Lautheitsgraden* verbunden werden, wie z.B. *schnell leise* (ital. *subito piano*). Beispiele für allmähliche Lautheitsveränderungen sind *lauter werdend* (ital. *crescendo*) und *leiser werdend* (ital. *decrescendo*) (siehe [Hem01]).

3.2 Taxonomien

Taxonomien beinhalten Klassifikationen von Entitäten einer bestimmten Art (z.B. Musik), wobei die Einheiten in Gruppen oder Kategorien zusammengefasst werden. Sie dienen somit zur Einordnung der einzelnen Instanzen (z.B. ein Lied). Im musikalischen Umfeld werden besonders die Musikgenres zur Organisation genutzt. Es sind aber auch weitere Taxonomien möglich, z.B.

- eine Instrumententaxonomie (mit z.B. Streichinstrumenten, Blasinstrumenten, Synthesizer, usw.),
- eine Musik(-signal-)eigenschaftentaxonomie (auch Spezialisierungen wie z.B. eine Akkordtaxonomie oder ein Tonarttaxonomie)
- aber auch eine Taxonomie über Stimmungen bzw. Stimmungslagen.

Die Entstehung von Kategorien folgt dem psychologischen Grundsatz: Menschen formen Kategorien indem sie die Notwendigkeit und grundlegenden Bedingungen für eine jede Kategorie spezifizieren [MF06]. Dabei gilt im Wesentlichen der Ansatz von Eleanor Rosch (auch bekannt unter *Prototypentheorie*; siehe [Ros73]): Kategorien werden hierarchisch organisiert und durch prototypische Exemplare definiert. Von diesen Erkenntnissen ausgehend hat Deliege folgende These für den Musikbereich formuliert (siehe [Del01]):

Menschen abstrahieren nützliche Merkmale, welche in Musikausschnitten enthalten sind und nutzen diese um Musik zu unterteilen, über Musikähnlichkeit zu urteilen und um *musikalische Fingerabdrücke*²¹ zu formen, welche helfen Musikstrukturen wahrzunehmen, Ähnlichkeiten auszuwerten und Klassifizierungen vorzunehmen.

Neben der Kombination von emotionalen, sozialen und kulturellen Faktoren, bilden auch jeweils typische Ausprägungen der (im vorherigen Unterkapitel beschriebenen) Musikmerkmale – insbesondere Tonhöhe, Klangfarbe, Rhythmus, Melodie, Harmonie und *Musikvortragsbezeichnungen* – die Grundlage für die jeweilige Definition der Musikgenres (vgl. [SCBG08] und [Les06]). Dabei gibt es z.Z. zwei grundlegend verschiedene Ansätze,

²¹Hier im Sinne von Prototypen, d.h. signifikante Grundmuster

Expertentaxonomien und *Folksonomien*, die im Folgenden vorgestellt, beschrieben und bewertet werden, und ein dritter Ansatz der hoffentlich in naher Zukunft sich durchsetzen kann.

3.2.1 Expertentaxonomien

Expertentaxonomien sind starr festgelegte Vokabeldefinitionen von Musikfachleuten. Das Resultat ist eine Liste von Termen, die in einer fest organisierten Hierarchie vorkommen und deren Beziehungen untereinander. Sie basieren oft auf der Definition von Genres. So gibt es Oberkategorien (z.B. Rock, Soul), welche die vielen Zusammenhänge der Unterkategorien vereinen, und die Musikstile (Unterkategorien; z.B. *Dark Metal*, *Northern Soul*), welche sich schnell von anderen Ausprägungen abgrenzen. Diese Definitionen können sich auch über mehrere Ebenen erstrecken, z.B. beinhaltet Abbildung 3.21 eine Auszug aus der *MP3.com*-Taxonomie mit 3 Schichten.

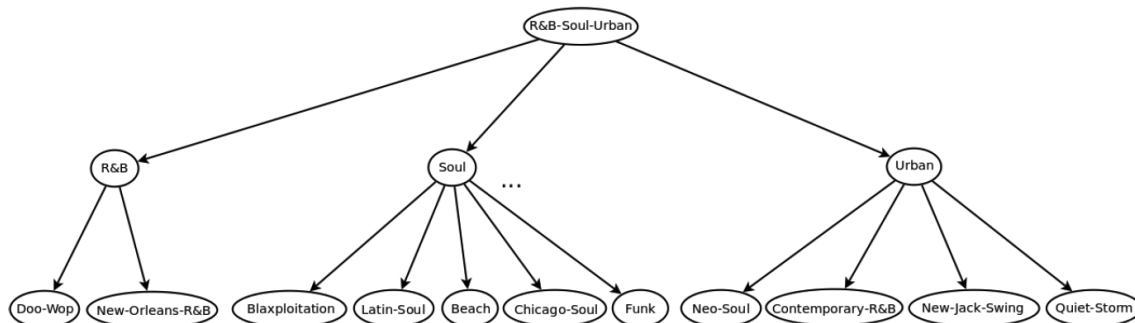


Abbildung 3.21: Ein Ausschnitt der Hierarchie aus der Taxonomie von *MP3.com* [SCBG08]

Der Nachteil dieses Ansatzes ist natürlich die Erhaltung der Aktualität dieser Definitionen. Auf der einen Seite kommen sehr schnell neue Kategorien und Unterkategorien hinzu (besonders im Bereich der elektronischen Musik). Auf der anderen Seite verändern sich manche Musikgenres auch ihr Grundzüge im Laufe der Zeit (z.B. Rock).

Generell haben diese Definitionen aber wenig Überschneidung. So ergab eine Untersuchung über verschiedene *Expertentaxonomie*-Definitionen (siehe F. Pachet und D. Cazaly [PC00]) nur eine Überschneidung von ca. 4,6 % (70 aus 1500 Genres). Zwei verschiedene Arten der *Expertentaxonomien* wurden nach Sordo et al.[SCBG08] definiert – die *Industrietaxonomien* und die *Internettaxonomien* – und unterschiedliche Eigenschaften aufweisen.

Industrietaxonomien Diese Taxonomiedefinitionen kommen meist von der Musikindustrie, z.B. Plattenfirmen (z.B. Virgin) oder Musikläden (z.B. Virgin Megastores), und sind daher auch kommerziell orientiert. Sie dienen der Kundennavigation und haben verschiedene Hierarchieebenen. Eine häufige Untergliederung ist zum Beispiel [SCBG08]:

- globale Musikkategorien
- Unterkategorien
- Künstler (alphabetisch geordnet)

- ggf. Alben

Eng verknüpft mit der Musikindustrie, sind die Gesellschaften zur Wahrung der Urheberrechte (z.B. GEMA²², ASCAP²³ oder SACEM²⁴), welche oft eine sehr grobe Einteilung haben und deshalb oft sehr fragwürdige Zuordnungen vornehmen (z.B. den Titanic-Soundtrack als Klassik, da die Plattenfirma oft klassische Musik vertreibt [PC00]).

Internettaxonomien *Internettaxonomien* (z.B. Amazon²⁵, *All Music Guide* (welche u.a. neben den Musikgenre auch die Stimmung zuordnen) oder *MP3.com*²⁶) sind den *Industrietaxonomien* ähnlich, d.h. sie sind im Grunde auch kommerziell orientiert. Zusätzlich bieten sie aber noch zahlreiche Querverweise zwischen den Künstlern, Alben etc. an und ermöglichen somit eine bessere Navigation für den Kunden und eine bessere Personalisierung des zugrunde liegenden Kataloges.

3.2.2 Folksonomien

Folksonomien entstehen durch das Zusammenwirken einer Gemeinschaft (*Community*) und bilden das Ergebnis des durch *Web 2.0* populär gewordenen *Social Taggings*, d.h. Benutzer beschreiben Inhalte durch Schlagwörter (*Tags*). Die Menge der Kennzeichner ist teilweise vorgegeben aber in der Regel auch beliebig erweiterbar, so dass jeder Benutzer seine individuellen Beschreibungen und Definitionen mit einfließen lassen kann. Sie folgen einem weniger formalen *Bottom-Up*-Ansatz bei welchem die Terme (Vokabeln) manuell durch den Annotationsprozess vom Benutzer entstehen. Somit können die *Folksonomien* auch sehr schnell wachsen und ggf. schwer handhabbar werden.

Ein populäres Beispiel einer *Folksonomie* im musikalischen Bereich ist *Last.fm* wobei die Benutzer persönliche Bezeichnungen eintragen können um die Musikstücke ihrer Musiksammlung so zu beschreiben und diese auch in Beziehung zu setzen. Generell erzeugt diese Möglichkeit eine höhere Benutzerzufriedenheit, da z.B. die Querverweise schneller zum Erfolg bei einer Suche führen können aber auch die verwendeten Begrifflichkeiten besser auf die Bedürfnisse der Gemeinschaft angepasst sind, weil sie von ihr definiert wurden. Ein Nachteil ist der nicht hierarchische Ansatz, bei welchem es Schwierigkeiten beim Einpflegen dieser oft sehr subjektiven Definitionen in Expertentaxonomien gibt.

Einer Untersuchung von Sordo et al. ergab, dass 39% der Annotationen von *Last.fm* mit den zugrunde liegenden Musikgenres der Expertentaxonomie übereinstimmten [SCBG08]. Die restlichen Kennzeichnungen dabei sind eines anderem Ursprungs (z.B. Instrumente, Stimmungen, andere Musikgenres oder Orte). Der in diesem Aufsatz vorgestellte Ansatz beschreibt ein Verfahren zur Abbildung von *Folksonomien* auf *Expertentaxonomien*, zeigt aber gleichzeitig auch auf, dass Musikgenretaxonomien nur ein Teilbereich des von den Benutzern definierten Vokabulars sind.

²²*Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte*; Verwertungsgesellschaft mit Sitz in Deutschland

²³*American Society of Composers, Authors, and Publishers*; Verwertungsgesellschaft mit Sitz in der USA

²⁴*Société des Auteurs, Compositeurs et Éditeurs de Musique* ; Verwertungsgesellschaft mit Sitz in Frankreich

²⁵<http://www.amazon.com/>

²⁶<http://www.mp3.com/>

3.2.3 Ein dritter Ansatz

Die vorgestellten Ansätze *Expertentaxonomie* und *Folksonomie* finden zwar in jeweils bestimmten Bereichen ihren guten Verwendungszweck, zeigen aber auch gleichzeitig gewissen Defizite auf, welche es gilt zu beseitigen. Im Allgemeinen haben die Musik-Genres für eine jede Person jedoch eine unterschiedliche Bedeutung in Abhängigkeit von der Herkunft und der Gruppierung dieser bzw. werden auch von ihr (der Person) neu definiert [SCBG08]. So besteht also immer noch die Schwierigkeit und zugleich die Herausforderung eine, im Übereinkommen von möglichst sehr vielen Menschen, gestaltete Musikategorisierung mit mit möglichst allgemeingültigen Definitionen der einzelnen Musikgenre und Musikstile zu finden.

An dieser Aufgabe sind schon viele Menschen gescheitert (siehe [Les06]), so kann es auch nicht Ziel dieser Arbeit sein eine solche Musikgenretaxonomie zu erstellen, sondern lediglich Empfehlungen zu geben und Informationsquellen und Repräsentationsformen zu beschreiben (siehe Kapitel 4 und 5), welche eventuell zu einer, für den individuellen Benutzer zufriedenstellen, Klassifikation oder (ggf. unbenannten) Gruppierung führen. Das Selbe gilt übrigens auch für die Definition einer Stimmungstaxonomie, welche i.A. noch ungenauere und subjektive Sichtweisen beinhaltet und damit die Zuordnung eines Musikstückes zu einer bestimmten Stimmungslage noch schwieriger gestaltet. So kann ein Mensch z.B. ruhige Rockballaden hören, wenn er traurig ist, und der Andere *Heavy Metal*. Demnach liegt der allgemeine Lösungsansatz in der Verarbeitung der individuellen Ansichten eines Benutzers, welche auf in (teil-)automatisierte Klassifikation oder Gruppierung mit verarbeitet werden. So kann z.B. eine auf unterschiedlichen Musikmerkmalen, -metadaten oder beschreibungen²⁷ festgestellte Anhäufung von Musikstücken (wenn die Analyse auf Basis von diesen durchgeführt wurde) entweder direkt einer vorher festgelegten Klassifikation zugeordnet werden, oder der noch nicht näher benannten Gruppierung durch einen (oder mehrere) Benutzer eine Bezeichnung gegeben werden.

Deshalb könnte vielleicht in naher Zukunft auch der Traum des dritten wesentlichen Ansatzes in Erfüllung gehen, welcher auf (teil-)automatischer Genreklassifikation beruht, die neben gewonnenen Informationen aus den Audiosignaleigenschaften, auch das Wissen der Expertentaxonomien und Folksonomien mit einfließen lässt und sich an den jeweiligen Benutzer anpasst (siehe Kapitel 5).

²⁷ Als eine Berechnung aus verschiedenen Audiosignaleigenschaften

4 Musikmetadaten

Abstrakte Beschreibungen von Musik, ihren Zusammenhängen und relevanten Konzepten können durch *Musikmetadaten* in einem bestimmten Datenmodell – ein *Musikmetadatenformat* – formuliert werden. Im Laufe der Jahre haben sich viele unterschiedliche Beschreibungsformen für diese Informationen entwickelt und etabliert. So gibt es nicht nur *formatgebundene Spezifikationen* (siehe Unterkapitel 4.2.1), wie z.B. *ID3*-Kennzeichner in *MP3*-Musikdokumenten, sondern auch *formatunabhängige Spezifikationen* (siehe Unterkapitel 4.2.2), welche in einer unterschiedlichen aber auch vielfältigen Weise genutzt werden können, z.B. die *Music Ontology*¹.

Darüber hinaus haben sich diverse *Musikmetadatendienste* (siehe Unterkapitel 4.3), z.B. *Discogs*², zur Aufgabe gemacht, Benutzer, direkt oder in aufbereiteter Form durch zusätzliche Anwendungen, mit vielfältigen Musikbeschreibungen zu versorgen. Dabei benutzen die *Musikmetadatendienste* i.A. eigene Metadaten schemata, z.B. hat *MusicBrainz* das *MusicBrainz XML Metadata Format (MMD)*, welche bei der Aufbereitung ggf. auf die verschiedenen *Musikmetadatenstandards und -ontologien* abgebildet werden.

Aus der Vielzahl der unterschiedlichen vorgestellten *Musikmetadatenformate* und unter Beachtung der darin enthaltenen *Musikmetadaten* können verschiedene Charakteristiken abgeleitet werden, welche ein *Musikmetadaten schema* zur Verwaltung von persönlichen Musiksammlungen haben soll (siehe Unterkapitel 4.4). Vor der Betrachtung der reichhaltigen Spezifikationen werden im folgenden Unterkapitel zu erst die Musikmetadaten auf unterschiedliche Weise kategorisiert.

4.1 Einordnungen von Musikmetadaten

Musikmetadaten können auf unterschiedliche Weise unterschiedlich differenziert eingeordnet werden. Die im folgenden Paragraphen vorgestellten *Metadaten typen* teilen dabei die *Musikmetadaten* in grobe Bereiche ein, wohingegen die *Metadatenfeldergruppierungen* eine eher feinere Separation anhand der eingeführten Anwendungsbereiche vornehmen.

Metadaten typen *Metadaten typen* sind Kategorisierungen von Metadaten, um diese einzuordnen und besser weiterverarbeiten zu können. Hier sollen kurz zwei Typisierungen vorgestellt werden, die sich von den Absichten her nicht überschneiden, aber die Metadaten doch wesentlich unterteilen und somit helfen die Qualität der Metadaten besser einzuschätzen.

Auf der einen Seite gibt es die allgemeine Unterteilung von Metadaten nach Duval et al., wobei in objektive und subjektive *Metadatenfelder* unterschieden wird (siehe [DHSW02]).

¹<http://www.musicontology.com/>

²<http://www.discogs.com/>

Angewandt auf die Musik sind objektive Elemente z.B. Künstler, Titel, Veröffentlichungsjahr und Version eines Musikstücktitels, d.h. Fakten die im Grunde unveränderbar sind und somit auch leicht automatisch ausgelesen werden können³. Es gibt aber auch eine Reihe subjektiver Eigenschaften, die durch persönliche Sichtweisen geleitet werden, z.B. Bekanntheitsgrad, Genre oder Stimmung. Generell kann somit jeder Parameter einer dieser beiden Kategorien zugeordnet werden und zeigt dadurch das Maß für die Gebrauchstauglichkeit der Werte an.

Auf der anderen Seite hat Pachet[Pac05] eine Unterteilung von *Musikmetadaten* in drei Kategorien vorgenommen, welche sich auf die Herkunft der Metadaten beziehen. In der ersten Gruppierung werden *redaktionelle Metadaten* zusammengefasst, d.h. Informationen, welche i.A. von einer Person, oft einem professionellen Redakteur oder Experten, manuell eingegeben werden. Darunter werden objektive wie auch subjektive Elemente erfasst, welche einen weiten Bereich abdecken – von administrativen Daten (u.a. auch Medium oder Datenformat⁴) über Entstehungs-, Aufnahme- und Aufführungsbeschreibungen bis zu historischen Fakten (z.B. Künstlerbiografien) [Pac05].

Neben den professionellen Kommentierungen (z.B. bei *AMG*) gibt es die Form der gemeinschaftlichen Bearbeitung durch eine Anwendergemeinschaft (z.B. bei *MusicBrainz*). Beide Arten sind durch ihrer Herkunft unterschiedlich kritisch in Bezug auf die jeweilige Subjektivität zu bewerten, welche durch inkrementelle Bewertungen und Resonanzen zu einer allgemeingültigeren Sicht führen kann. Insbesondere gilt dies auch bei der Formulierung von Taxonomien (siehe Abschnitt 3.2), wo es u.a. die Aufgabe ist *Expertentaxonomien* unterschiedlicher Bereiche (z.B. Musikgenre oder Stimmung) mit *Folksonomien* zusammenzuführen, um daraus für den Benutzer zufriedener stellende Klassifikationen zu bekommen, welche sich auch der entwickelnden Natur der Musik anpassen.

In der zweiten Kategorie werden *kulturelle Metadaten* zusammengefasst, welche durch ein gewisses Milieu oder eine Kultur entstehen und somit der Musik einen gewissen Kontext geben. Sie sind das Ergebnis der Analyse von vorkommenden Mustern, Kategorien und Assoziationen einer Quelle von Dokumenten bzw. Daten[Pac05] (siehe Abschnitt 5.2.2). Eine geläufige Methode ist dabei das *Gemeinschaftliche Filtern*⁵ mit einer Menge von Benutzerprofilen als Quelle, die oft für Empfehlungen genutzt werden (z.B. bei *Amazon* oder *Last.fm*). Weitere Quellen sind Websuchmaschinen, Musikradioprogramme oder einfache Textquellen wie Bücher. Die wesentlichen Techniken kommen dabei von der Computerlinguistik, deren Gegenstand die maschinelle Verarbeitung natürlicher Sprache⁶ ist, und beruhen weitestgehend auf Übereinstimmungsanalysen⁷, bei welchen bedeutungsvolle Ähnlichkeitsbeziehungen (z.B. Ähnlichkeitsdistanzen oder Wortzusammenhänge) aus den Übereinstimmungsinformationen abgeleitet werden (siehe auch Abschnitt 5.2.3).

Die letzte Kategorie von Pachet enthält *akustische Metadaten*, welche vollständig aus der Analyse des Audiosignales abgeleitet werden. Sie versuchen eine gewisse Objektivität zu vermitteln, obwohl ein Musikstück relativ unterschiedlich von einem Hörer wahrge-

³Von der Manipulation oder Anpassung gewisser Kennzeichner wird hierbei abgesehen

⁴Es gibt auch Ansätze, die diese Informationen in eine separate Kategorie, den *physikalischen Metadaten* abgrenzen (siehe [DP06]). Da diese Kategorie sich aber nicht auf der gleichen Ebene bewegt wie die anderen, wird sie mit in die wesentlichen Metadaten, den *redaktionellen Eigenschaften*, eingeordnet.

⁵engl. *collaborative filtering*; siehe <http://www.adastral.ucl.ac.uk/~junwang/CollaborativeFiltering.html>

⁶engl. *Natural Language Processing, NLP*

⁷engl. *co-occurrence analysis*; siehe <http://www.similarity-blog.de/publications/>

nommen werden kann und es deshalb eine Vielzahl an subjektiven Eigenschaften besitzt (siehe Kapitel 3). In den *akustischen Metadaten* werden alle Dimensionen der Musikwahrnehmung abgedeckt und die Informationen reichen von einfachen Tempoangaben über Rhythmusstrukturen hin zu Perkussivität und Instrumentenerkennung. Neben diesen unären Beschreibungen zählen auch nicht-unäre Beschreibungen wie Melodiekonturen oder Tonhöhenenerkennung zu diesen *Musikmetadatentyp*.

Teilweise können gewisse Eigenschaften, z.B. Musikgenre, auch in jeder dieser Kategorien ihren Ursprung haben (siehe Abschnitt 3.2.3) und deshalb auch miteinander verglichen, vereint oder aufeinander aufbauend genutzt werden (z.B. als Trainingsdatensatz für *Maschinelles Lernen*, siehe Abschnitt 5.2.3). Durch die beschriebene Typisierung von *Musikmetadaten* können alle folgenden *Metadatenfelder* der jeweiligen Spezifikation gut eingeordnet und für die jeweiligen speziellen Anwendungsbereiche wiederverwendet werden.

Anwendungsbereiche Die vorgestellten Musikmerkmale, Musikmetadatendienste, -standards und -ontologien werden in unterschiedlichen Anwendungsbereichen genutzt bzw. sind dafür geeignet. Diese allgemeinen Anwendungsbereiche basieren generell auf den Handlungen, die Menschen ausführen, wenn sie sich mit Musik auseinandersetzen. Die wesentlichen Umgebungen, welche durch Software abgedeckt werden können, werden im Folgenden kurz erläutert (nach [CGVD08]):

- **Musikbibliotheken und -enzyklopädien:** für Bibliotheken, Enzyklopädien, Musiklizenzierungsfirmen oder Unternehmen, welche sich mit der Faktenbeschreibung von großen Musiksammlungen beschäftigen (z.B. *Macrovision* mit dem *All Music Guide*)
- **Verwaltung von persönlichen Musiksammlungen:** zum Organisieren von eigenen Musiksammlungen (z.B. *iTunes*, *WinAmp*⁸)
- **Handel und Geschäft:** Anwendungen, welche den Erwerb von Musik, egal welchen Formates, ermöglichen (z.B. *iTunes*, *Amazon*)
- **Musikbearbeitung und -produktion:** für Musikeigenkreationen und -weiterbearbeitungen (z.B. *Logic Pro*⁹)
- **Musikwiedergabe:** Anwendungen zur Wiedergabe von Musikdokumenten (z.B. *WinAmp*)
- **Musikempfehlung:** Dienste zum Entdecken von neuer oder ähnlicher Musik (z.B. *Pandora*¹⁰, *Last.fm*)
- **Musikinformationsretrieval:** Such- und Erkennungswerkzeuge für unterschiedliche Anfrageschnittstellen (z.B. Anfragen durch Summen)
- **Musiknotation:** Anwendungen zum Erstellen und Verändern von musikalischen Partituren (z.B. *Sibelius*¹¹).

⁸<http://de.winamp.com/>

⁹<http://www.apple.com/de/logicstudio/logicpro/>

¹⁰<http://www.pandora.com/>

¹¹<http://www.sibelius.at/>

Reale Anwendungen beinhalten oft mehrere Anwendungsbereiche, z.B. kann eine Anwendung, welche Wiedergabelisten erzeugt der Musikempfehlung und der Bibliotheksfunktionalität zugeordnet werden. Welche dieser Kategorien zur Handhabung von Musik im privaten Umfeld zum Einsatz kommen, erläutert Kapitel 2 detailliert.

Metadatenfeldergruppierungen Jede Spezifikation besitzt ihre eigenen, verschiedenartig benannten *Metadatenfelder*. Um einen Vergleich der Ausdruckskraft und Reichhaltigkeit bzw. Wertigkeit eines Metadatenschemas für einen Anwendungsbereich zu ermöglichen, wurde in der Arbeit von Corthaut et al. [CGVD08] eine Definition von Gruppierungen semantisch zusammenhängender *Metadatenfelder* im Bereich der Musikmetadatenstandards vorgenommen. Mit diesen *Metadatenfeldergruppierungen* ist es auch möglich, die unterschiedlichen Granularitätsebenen der Beschreibungen zu beherrschen. Im Folgenden sollen diese semantisch verwandten Elemente, welche unabhängig von Metadaten schemata und deren *Metadatenfeldern* sind, vorgestellt werden (nach [CGVD08]):

- **Musikalische Informationen:** Eigenschaften des Audiosignals (z.B. BPM, Länge des Musikstückes)
- **Klassifizierungen:** Kategorien für die Musikorganisation z.B. Genre oder Stimmung
- **Aufführung:** Beschreibung einer Darbietung (oft verbunden mit einer Aufnahme) und der Menschen, die an diesem Prozess beteiligt sind (z.B. Künstler, Dirigenten, Techniker)
- **Versionsverwaltung:** Beschreibung von verschiedenen Versionen eines Musikwerkes und die Beziehung der beteiligten Menschen und Werke, z.B. Eigenschaften wie *ursprünglicher Urheber; enthält Ausschnitte von; Remix von*
- **Musikwerksbeschreibungen:** Beschreibung des musikalischen Werkes in seiner abstrakten Form, z.B. Musikstücktitel
- **Rechte und Urheber:** Informationen zur Verwaltung von geistigem Eigentum, z.B. Lizenz, Urheber/ Rechteinhaber
- **Wiedergabe:** Informationen zum Wiedergeben von Musikdokumenten, z.B. relative Lautstärkeanpassung
- **Liedtext:** Text des Musikwerkes und verwandte Informationen, z.B. übersetzte Liedtexte, abgewandelte Liedtexte
- **Einordnung und Referenzierung:** Dateneinordnungsstrukturen und Verweise auf Webseiten und anderen Werke, z.B. Wikipedia-Einträge, Alben-Auflistungen
- **Identifikatoren:** Identifikationsschlüssel, z.B. ISRC-Kennung¹², Katalognummer einer Veröffentlichung
- **Aufnahme:** Informationen über die Aufnahme und den daran beteiligten Personen, z.B. Aufnahmetyp, Albumtitel

¹²*International Standard Recording Code*

- **Instrumentation und Anordnung:** Informationen über die genutzten Instrumente und Orchestrierung, z.B. Instrumententyp
- **Klang- und Signalträger:** Informationen über verfügbare Medien eines Musikwerkes, z.B. Medientyp
- **Anlass:** Informationen über öffentliche Darbietungen eines Musikwerkes, z.B. Festivalprogramm, Ort
- **Zeitmodellierung:** Konzepte zum Modellieren von zeitlichen Informationen, z.B. Intervall, Dauer
- **Musiknotation:** für die symbolische Repräsentation von Musik, z.B. Notenschlüssel, Takt, Noten
- **Beachtung und Gebrauch:** Informationen über die Beachtung eines Musikstückes durch Benutzer, z.B. Anzahl der Wiedergaben
- **Veröffentlichung:** Informationen über den Veröffentlichungsvorgang, z.B. Plattenfirma
- **Komposition:** Informationen über den Kompositions- bzw. Entstehungsprozess, z.B. Satz, Komponist
- **Produktion:** Informationen über den kreativen Prozess in der Produktion und deren Beteiligte, z.B. Produzent
- **Meta-Metadaten:** Informationen über die Metadaten, z.B. Kommentator der Metadaten

Je nach Anwendungsbereich ergibt sich eine unterschiedliche Relevanz für die verschiedenen *Metadatenfeldergruppierungen*, z.B. sind bei der Wiedergabe von Musikstücken *Musikwerksbeschreibungen* von höherer Bedeutung als Informationen über die Aufnahme (obwohl die ein Benutzer ggf. auch repräsentiert haben möchte). Die Bedeutung der genannten Elemente soll im anschließenden Unterkapitel detailliert für die Verwaltung von persönlichen Musiksammlungen untersucht werden.

4.2 Musikmetadatenstandards und -ontologien

Musikmetadatenstandards und Ontologiedefinitionen dienen der zusätzlichen Beschreibung von Musikstücken und helfen somit die persönliche Musiksammlung besser zu organisieren und besser in ihr zu navigieren. Grundsätzlich lassen sich zwei Spezifikationsarten unterscheiden. Die durch die Entwicklung von Audiosignalformaten entstandenen *formatgebundenen Spezifikationen*, und die allgemeineren *formatunabhängigen* Metadatenstandards und Ontologiedefinitionen.

Der Ursprung bzw. der Entstehungsprozess der Metadatenstandards ist unterschiedlich. Als Erstes kann sich aus der Entwicklung einer Anwendung heraus ein De-Facto-Standard entwickeln. Als Zweites kann die Gestaltung mit dem Ziel auf Interoperabilität erfolgen, d.h. *formatunabhängige Spezifikationen*, welche *formatgebundene Spezifikationen* mit Metadaten versorgen. Als Drittes kann eine Metadatenbeschreibung auch von Beginn

an als Standard vorgesehen sein, z.B. *MPEG-7* [CGVD08].

Allgemein gilt der Grundsatz: *formatunabhängige Spezifikationen* versuchen die Schwachstellen der einzelnen *formatgebundenen* Beschreibungsformate zu kompensieren und darüber hinaus ein höhere Maß an Interoperabilität zu erzielen. Es existiert z.Z. aber immer noch kein Musikmetadatenstandard, welcher alle Anforderungen der verschiedenen Anwendungsbereiche auf einmal erfüllt [CGVD08].

4.2.1 Formatgebundene Spezifikationen

Formatgebundene Spezifikationen sind im Anwendungsbereich der Musikwiedergabe weit verbreitet. Dabei werden im Allgemeinen Datencontainer in einem vorgeschriebenen Format mit in eine Audio-/Multimediadatei (des jeweiligen Audio-/Multimediadateiformates) eingebunden (am Anfang oder am Ende).

ID3 Das Beschreibungsformat des am häufigsten genutzten Audiosignaldateiformates *MP3* heißt *ID3*-Kennzeichner. Die erste Version (*ID3v1*) enthält 5 Metadatenfelder (Musikstücktitel, Künstler, Album, Jahr, Kommentar und Genre) fester Länge, welche bei der Nachfolgevariante (*ID3v1.1*) variabel gestaltet wurde. Zusätzlich enthält diese Abwandlung ein weiteres *Metadatenfeld* (Albumtitelnummer)[O’N06a].

Eine grundlegende Erweiterung dieses Beschreibungsformates ist die Version *ID3v2*, welche sich z.Z. in der Entwicklungsstufe *ID3v2.4* befindet. Trotzdem ist *ID3v2.3* die am weitesten eingesetzte Variante, d.h. ein inoffizieller De-facto Standard[O’N09]. Die Gestaltungsziele für die Version *ID3v2* sind Flexibilität und Erweiterbarkeit, welche sich in einer großen Anzahl neuer *Metadatenfelder* (u.a. Darsteller, Webseite, Liedtext, Equalizer-Voreinstellungen, Komponist, Urheberrechtsvermerk, genutzte Sprache, Wiedergabezähler, Popularitätsrate¹³) widerspiegeln. Darüber hinaus gibt es auch die Möglichkeit eigene *Metadatenfelder* hinzuzufügen [O’N06b].

Weiterhin wird von dieser Version an Unicode, die Verschlüsselung und Verknüpfung (Weblinks) von Informationen, die Einbindung von Bildern und die Anbindung an CD-Datenbanken (*CDDB* u.a.) unterstützt. *ID3v2* ist zwar grundlegend für die Formate *MPEG-1* u. *MPEG-2 Audio Layer I - III* gestaltet, kann aber auch für andere Zielformate verwendet werden [O’N06b].

Im Allgemeinen haben die *ID3*-Kennzeichner den großen Nachteil, dass sie nur aus einfachen Bezeichner-Wert-Paaren bestehen und somit nicht flexibel genug sind mehrdeutige Eigenschaften gut auszudrücken, d.h. es kann z.B. immer nur ein Musikgenre angegeben werden. Die direkte Einbindung von Bildmaterial, z.B. Coverbilder, ist jedoch ein positives Kriterium, da Benutzer i.A. eine visuelle Repräsentation eines Musikstückes oder einer Musikzusammenstellung bevorzugen, da sie es aus ihrer herkömmlichen (analogen) Musiksammlung so gewöhnt sind (vgl. [CJJ04]).

Vorbis Comments Dieses Metadatenformat wurde ursprünglich für das Audiosignaldateiformat *Ogg Vorbis* entworfen, wurde aber u.a. auch für weitere Audio-Codecs (*FLAC*, *Speex*) von *Xiph.Org*¹⁴ verwendet. Die Definition erfolgt ähnlich wie bei *ID3*-Kennzeich-

¹³Eine genaue Auflistung der *Metadatenfelder* ist unter <http://www.id3.org/Frames> zu finden

¹⁴<http://www.xiph.org/>

nern mit Bezeichner-Wert-Paaren, im Gegensatz zu diesem Metadatenformat ist bei *Vorbis Comments* keine feste Struktur vorgegeben. Die Ersteller geben lediglich Empfehlungen, welche *Metadatenfelder* (Bezeichner) verwendet werden sollten [Xip09]. Durch diesen offenen Ansatz können die *Metadatenfelder* auch mehrfach verwendet werden, was z.B. eine vielfache Kennzeichnung mit Musikgenres ermöglicht [Xip05].

Da dieses Beschreibungsformat keine vorgeschriebene *Metadatenfelddenutzung* hat, existieren etliche Erweiterungsansätze¹⁵ aus der Entwickler-Community, welche eine Vielzahl weiterer Informationen bereithalten, u.a. eine *Metadatenfeldergruppierung* für die beteiligten Personen und Urheberrechtsinformationen. Die Einbindung von Bildern (wie bei *ID3v2*) ist offiziell leider nicht möglich. Hierbei existieren wiederum auch verschiedene Ansätze, die aber zu Inkonsistenzen bei den *APIs* und Anwendungen führen, die dieses Metadatenformat verarbeiten [Xip09]. Ähnlich wie bei *ID3v2*-Kennzeichnern, wird der Unicode-Zeichensatz für den Zeichenketten der Werte unterstützt.

Eine Weiterentwicklung für die Beschreibung von Metadaten für *Ogg Vorbis* ist das *XML*-basierte *Multimedia Metadata Format (M3F)*, welches *Vorbis Comments* ersetzen soll und eine Vielzahl von Feldern definiert, die optional belegt werden können. Dieser Ansatz befindet sich aber noch in einem frühem Stadium der Entwicklung [Xip08].

APE Ein weniger bekanntes Metadatenformat, welches anfänglich für das Audiosignaldatenformat *Monkey's Audio*¹⁶ entwickelt wurde, ist *APE*. Mittlerweile existiert es in der Version *APEv2*, welche für den Audio-Codec *Musepack (MPC)*¹⁷ gestaltet wurde, aber auch bei *WavPack*, *OptimFROG* und weiterhin auch bei *Monkey's Audio* genutzt wird [mis08b]. Das Beschreibungsformat ist ähnlich zu *Vorbis Comments* angelegt.

Die wesentlichen Unterschiede zwischen *APEv1* und *APEv2* sind die Unterstützung von Unicode (anstatt ASCII) und ein *Header*-Teil am Anfang der *APEv2*-Spezifikation (siehe [mis08a]). Ähnlich dem Musikmetadatenformat *Vorbis Comments* bieten *APE*-Kennzeichner die Möglichkeit, mehrere Werte für einen Bezeichner zu definieren, nur dass sie hierbei in einer Liste angelegt werden und nicht als separates Bezeichner-Wert-Paar [mis08b].

4.2.2 Formatunabhängige Spezifikationen

Formatunabhängige Spezifikationen sind im Anwendungsbereich der Musiksuche und Musikempfehlung weit verbreitet, finden aber auch in der Musikwiedergabe und der Sicherung von Audio-CDs Verwendung. Sie beinhalten im Allgemeinen eine Vielzahl von Audiosignaleigenschaften in aufbereiteter Form (z.B. Klangfarbe, *Beats per Minute (BPM)* oder Takt) und zusätzlichen Beschreibungen (z.B. über den Entstehungsprozeß), um den Anforderungen der genannten Einsatzgebiete gerecht zu werden.

Die Spezifikationen nutzen dabei unterschiedliche Datenmodelle und Beschreibungsformate, z.B. *RDF* und *N3*[BLCK⁺08], oder *XML* und *Turtle*¹⁸, zur Definition der Standards und Ontologien, und zur Repräsentation ihrer Instanzen. Ontologiebasierte Repräsentationen werden im Weiteren deshalb oft als *semantische Graphen* bezeichnet – unabhängig vom genutzten Datenmodell.

¹⁵z.B. dieser <http://reallylongword.org/vorbiscomment/>

¹⁶vergleichbar zu *FLAC*

¹⁷basierend auf den *MPEG-Layer-2*-Algorithmen

¹⁸*Terse RDF Triple Language*, siehe <http://www.w3.org/TeamSubmission/turtle/>

4.2.2.1 MPEG-Standards

Die verschiedene Datei-, Inhalts- und Metadatenformate der *MPEG (Moving Picture Experts Group)*¹⁹⁾ können auch im Bereich der Musikdokumente und Musiksammlungen eingesetzt werden. Zur Beschreibung von Musikmerkmalen und -metadaten können der *MPEG-7*- und der *MPEG-21*-Standard verwendet werden, welche im Folgenden vorgestellt werden.

MPEG-7 Dieser *Multimedia*-Inhaltsbeschreibungsstandard²⁰⁾ bietet genormte Schemata zum Beschreiben von verschiedenen *Multimedia*-Aspekten auf verschiedenen Abstraktionsebenen und ergänzt somit die bestehenden *MPEG-1*-, *MPEG-2*- und *MPEG-4*-Spezifikationen, welche weitestgehend nur der Repräsentation von Inhalten dienen. Ein primäres Entwurfsziel bei der Gestaltung der *MPEG-7*-Spezifikation war es, eine hohe Interoperabilität zu erreichen. Dabei wurden die folgenden Gestaltungsprinzipien angewandt:

- möglichst viele Anwendungsbereich abdecken
- verschiedene Beziehungen zum Inhalt herstellen (als separate Beschreibung, in Verbindung mit dem Inhalt, verknüpft mit dem Inhalt)
- eine breite Unterstützung verschiedenster Datentypen ermöglichen (für Sprache, Audio, Bilder, Video, Grafik, 3D-Modelle, ...)
- eine Unabhängigkeit zu den Audiosignalträgern zu schaffen, welche das *Multimedia*-Signal enthalten
- objektbasiert
- Formatunabhängigkeit
- verschiedene Abstraktionsebenen
- Erweiterbarkeit

Die Definition der Beschreibungen und Beschreibungsschemata erfolgt mittels eines erweiterten *XML*-Schemas, der *Description Definition Language (DDL)*, welche wiederum auch zur Erweiterung der vorhandenen Spezifikation genutzt werden kann. Die Informationen können entweder textuell (*TeM*²¹⁾), als *XML*-Instanz-Dokument des *MPEG-7*-Schemas für besseren Zugang zum Editieren, Suchen und Filtern, oder binär (*BiM*²²⁾), als kompaktes Format zum Speichern, Übertragen und Ausliefern, vorliegen.

Es gibt allgemeine *Multimedia*-Beschreibungsschemata (*MDS*) und anwendungsspezifische Beschreibungsschemata, u.a. eines für den Bereich Audio. Deshalb wird auch im hier behandelten Anwendungsbereich oft die Bezeichnung *MPEG-7 Audio* verwendet. Da bei der Gestaltung der *MPEG-7*-Spezifikation nach dem „one functionality, one tool“-Prinzip vorgegangen wurde (um Redundanzen zu vermeiden), muss eine Implementierung/ Anwendung im Allgemeinen mehrere Beschreibungsschemata nutzen, um ihren jeweiligen

¹⁹<http://www.chiariglione.org/mpeg/>

²⁰Multimedia Content Description Interface; ISO/IEC 15938

²¹„Textual format for Mpeg-7“[MSS03]

²²„Binary format for Mpeg-7“[MSS03]

Anwendungsbereich abzudecken [MSS03].

Auf der einen Seite decken die allgemeinen Beschreibungsschemata unveränderliche Metadaten in Verbindung zum Erstellungsprozess, Produktionsprozess, der Nutzung und der Verwaltung der Multimedia-Inhalte ab, und dienen zur direkten Beschreibung des Inhaltes auf verschiedenen Ebenen (Signalstruktur, Merkmale, Modelle und Semantiken). Auf der anderen Seite erschließen die anwendungsspezifischen Beschreibungsschemata Merkmale und Eigenschaften, die nur in diesen Bereichen vorkommen.

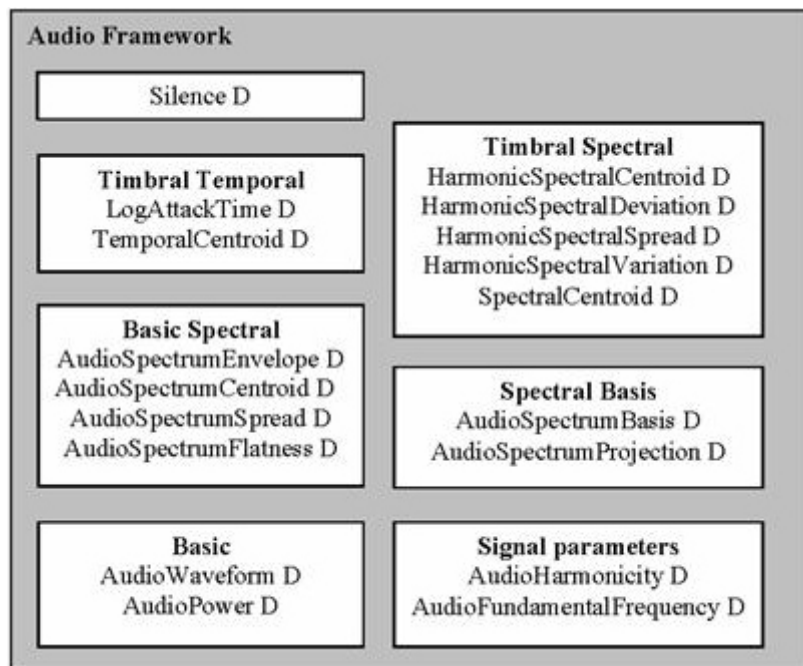


Abbildung 4.1: Eine Übersicht der *low-level*-Beschreibungen von *MPEG-7 Audio*[Mar04]

Die Audio-Beschreibungen (*MPEG-7 Audio*) unterteilen sich in zwei verschiedene Bereiche. Als Erstes dienen die *low-level*-Beschreibungen (*AudioLLDs*) zur standardisierten Beschreibung von Analyseergebnissen des Audiosignals (was in der Regel automatisiert erfolgt) und bieten somit auch eine Grundlage an Werten und Informationen für die Weiterverarbeitung. Die in der Audiosignalanalyse verwendeten Algorithmen werden dabei nicht vorgeschrieben, es existieren aber empfohlene Vorlagen.

In Abbildung 4.1 sind die möglichen Beschreibungen dieser Klasse aufgelistet, welche von grundlegenden Eigenschaften (z.B. Wellenform) über einfache Spektralanalysen hin zu Klangfarbenspektren reichen. Die zweite Klasse sind die *high-level*-Beschreibungen, welche nach ihren verschiedenen Anwendungsbereichen kategorisiert sind. So gibt es Beschreibungen für die allgemeine Klangerkennung und Indexierung, die Instrumentenklangfarbe, gesprochenen Inhalt, Audio-Signaturen und Melodiebeschreibung.

Mit Hilfe der Spektralanalyse (*Spectral Basis*) aus den *low-level*-Beschreibungen und den abstrakten Beschreibungsschemata für Klangmodell und Klangklassifikationsmodell²³ lassen sich Klangkategorien (z.B. Musikinstrumenten) und Ähnlichkeitsmodelle dazu erstellen. Diese können für die automatische Indexierung und Einordnung (z.B. Musikgenre)

²³abgeleitet aus den *MDS* für *Hidden-Markov-Modell (HMM)* und Wahrscheinlichkeitsklassifizierungsmodell; siehe Abschnitt 5.2.3

genutzt werden können.

Da der *MPEG-7*-Standard über einen Zeitraum von mehreren Jahren entwickelt wurde, gab es auch diverse Abänderungen im Laufe der Zeit. So wurden u.a. Audiobeschreibungen für Signalqualität, Tempo, Akkord und Rhythmus hinzugefügt, oder die allgemeinen Beschreibungsschemata um linguistische Beschreibungen erweitert [Mar04]. Trotzdem fehlen bei dieser Spezifikation immer noch genaue Semantikdefinitionen. Auf der einen Seite können semantisch identische Inhalte auf unterschiedlichen Wegen repräsentiert werden (Freitextannotationen, Schlagwortannotationen oder strukturierte Annotationen mit Kennzeichnern). Auf der anderen Seite sind die beabsichtigten semantischen Bedeutungen der beschriebenen Strukturen nicht immer eindeutig verständlich, z.B. wird bei einer Personensuche nicht gleichzeitig ein Bild, das mit dieser Person gekennzeichnet ist, und ein Bildausschnitt, welcher mit dieser Person gekennzeichnet ist, angezeigt) [CDH⁺07].

Deshalb existieren zwei Ansätze, von Hunter[Hun01, HL06b] und Tsinaraki et al.[TPC04], welche innerhalb eines *W3C*-Vorschlages die *MPEG-7*-Spezifikation auf eine Ontologie abbilden. Der erste Ansatz nutzt dabei die *ABC*-Ontologie²⁴ als Grundlage, wohingegen der Zweite eine eigene Ontologiedefinition verwendet. Darüber hinaus entstand noch im Rahmen der *Rhizomik*-Initiative²⁵ eine automatisch generierte *MPEG-7*-Ontologie mittels einer generischen *XSD*-zu-*OWL*-Abbildungstechnologie namens *XSD2OWL* (siehe [GsC05]). Dieser Ansatz ermöglicht u.a. die Umwandlung von *XML*-Instanz-Dokumenten des *MPEG-7*-Schemas in *RDF*-Instanz-Dokumente dieser *MPEG-7*-Ontologie.

MPEG-21 Neben dem vorgestellten Musikmetadatenformat *MPEG-7 Audio* zur inhaltlichen Beschreibung von *Multimedia*-dokumenten, gibt es noch ein weiteren Metadatenstandard der *Moving Picture Experts Group*, welcher im Umgang mit Musikedokumenten betrachtet werden muss – der *MPEG-21*-Standard. Dieses Spezifikation ermöglicht die transparente und vermehrte Nutzung von *Multimedia*-Ressourcen über eine Vielzahl von Netzwerken, Geräten, Nutzereinstellungen und Gemeinschaften, d.h. die Transaktionen von *digitalen Gegenständen (Digital Items)*, für den in dieser Arbeit beschriebenen Anwendungsfall enthalten sie Musikedokumente, zwischen *Benutzer (Users)*, d.h. in diesem Fall die Besitzer der persönlichen Musiksammlungen, Personen, welchen sie ihre Musikedokumente zugänglich machen oder ausleihen, aber auch Softwareanwendungen, die direkt mit der Anwendung zur Verwaltung der Musiksammlung in Verbindung treten (siehe [Tim05, Tim07]).

Diese Metadaten schemata beschreibt somit eine Infrastruktur zum, Ausliefern, Austauschen und Konsumieren von *Multimedia*-Inhalten. Dabei beinhaltet ein *Digital Item* die Ressourcen (hier Musikedokumente) an sich, zugehörige Metadaten und die Struktur zwischen den Ressourcen und Metadaten. Ein *User* kann dabei auf verschiedenste Weise mit den *Digital Items* interagieren, z.B. Ressourcen modifizieren, archivieren, bewerten, verteilen oder zusammenstellen, und hat dabei verschiedene spezielle Rechte und Verantwortlichkeiten in Bezug auf eine Interaktion [Tim05].

Für die Verwaltung von persönlichen Musiksammlungen, sind besonders die Beschreibungsmöglichkeiten zur *Anpassung der digitalen Gegenstände (Digital Item Adaption)* interessant, welche auch eine Modellierung der Gebrauchsumgebung ermöglichen. Somit können z.B. Benutzercharakteristiken mit Informationen über dem Anwender selbst, über

²⁴<http://metadata.net/harmony/ABC/ABC.owl>

²⁵<http://rhizomik.net/>

seine Voreinstellungen (z.B. Filter- und Sucheinstellungen) und seinen Benutzungsverlauf erstellt werden [TC07].

Der *MPEG-7*- und *MPEG-21*-Standard lassen sich gut miteinander kombinieren. Tsinaraki et al.[TC07] haben in ihrer Arbeit *MPEG7/21*-Rahmenwerk entworfen, welches zuerst die jeweiligen Schemata in eine *OWL*-Ontologie überführt und zusätzlich anwendungsspezifische Ontologien mit in die Beschreibungen integrieren kann, um z.B. ein besseres Filtern, Durchsuchen und Interagieren mit *Multimedia*-Datenbanken zu ermöglichen.

4.2.2.2 Das *Music Ontology*-Rahmenwerk

Als Bestandteil eines größeren, erweiterbaren Systems, welches zur Beschreibung und Verknüpfung von musikspezifischen Informationen innerhalb des *Semantic Web* dient, wurde die *Music Ontology* entworfen. Das große Ziel dieser Reihe von Ontologien ist es eine riesige, verteilte, maschinenlesbare kulturelle Umgebung zu schaffen, welche es schafft, alle Konzepte der Musikproduktion, des Musikgebrauchs, der Musikempfehlung usw. zu behandeln [RASG07]. Des Weiteren können auch verschiedene Wissenbasen (z.B. Online-Datenbanken oder Daten von Audiosignalextraktionsdiensten) als Grundlage genutzt werden und diese Inhalte effektiv vereint werden.

Für diesen Zweck müssen die musikspezifischen Informationen wohldefinierte Bedeutungen aufweisen, damit diese auch gut weiterverarbeitet und wieder benutzt werden können [Rai08a]. Durch die Benutzung einer speziellen Beschreibungslogik (*SHOIN(D)*²⁶) als Teilmenge der Prädikatenlogik erster Ordnung können die Zusammenhänge zwischen den Konzepten und Eigenschaften des Musikanwendungsbereiches genau definiert werden und bekommen gleichzeitig einen Rahmen, welcher auch die Grenzen bei Erweiterungen aufzeigen soll, damit die Ontologie nicht zu einem allumfassenden unentscheidbaren System wird [Rai08a].

Bei der Modellierung des *Music Ontology*-Rahmenwerkes werden Konzepte des *RDF*-Standards verwendet, d.h. die Beschreibungen erfolgen über Subjekt-Prädikat-Objekt-Tripel, zum Identifizieren dieser Quellen werden *URIs* genutzt und zur kompakten Repräsentation von *semantischen Graphen* werden *Named Graphs*²⁷ eingesetzt. Weiterhin wird die *SHOIN(D)*-Beschreibungslogik durch die Verwendung der *OWL*-Teilsprache *OWL DL* realisiert.

Bei der Entwicklung dieser Ontologie wurde großer Wert auf Wiederverwendung gelegt, d.h. es wurden möglichst viele Beschreibungsstandards mit eingebunden, u.a.

- das *Dublin-Core*-Schema²⁸ (*DC*), zur Beschreibung von Kalenderdaten und Titeln,
- die *Friend-of-a-Friend*-Ontologie²⁹ (*FOAF*), zum Modellieren von Personen und ihren Beziehungen, oder
- die *Functional-Requirements-for-Bibliographic-Records*-Ontologie³⁰ (*FRBR*), zur Ableitung grundlegender Konzepte, welche es ermöglichen Musikobjekte auf unterschiedlichen Ebenen zu modellieren (siehe Paragraph „*Music Ontology*“).

²⁶Wobei (D) die Eigenschaft des benutzten Datentyps ist; siehe <http://www.cs.man.ac.uk/~ezolin/dl/> für weitere Informationen

²⁷siehe <http://www.w3.org/2004/03/trix/>

²⁸<http://purl.org/dc/elements/1.1/>

²⁹<http://xmlns.com/foaf/0.1/>

³⁰<http://purl.org/vocab/frbr/core#>

Zusätzlich bereichern weitere Ontologien die Basis des *Music Ontology*-Rahmenwerkes, können aber durch ihre Eigenständigkeit und Allgemeingültigkeit auch in anderen spezifischen Anwendungsbereichen eingesetzt werden. Diese Kernontologien werden im Folgenden näher vorgestellt und danach die Hauptontologie des *Music Ontology*-Rahmenwerkes vorgestellt – die *Music Ontology*.

Timeline Ontology Auf der einen Seite existiert die *Timeline Ontology*³¹, welche auf *OWL-Time*-Ontologie³² aufbaut, die den Stand eines *W3C*-Standardentwurfs hat. Das wesentliche Konzept dieser Spezifikation, die Zeitleiste (*TimeLine*), baut dabei auf den Begriffen Intervall (*Interval*) und Moment (*Instant*) auf. So können z.B. Augenblicke in einem Audiosignal definiert werden (durch diskrete Zeitleisten bei digitalen Signalen) aber auch Zeitbereiche auf einem Kalender (durch fortlaufende Zeitleisten). Abbildung 4.2 illustriert diese beschriebenen Beziehungen und zeigt gleichzeitig auch die Vielfältigkeit dieser Ontologie zum Darstellen von Zeitinformationen. Der obere *semantische Graph* stellt einen Moment in einem Audiosignal dar (bei 3 Sekunden), und der untere *semantische Graph* ein Intervall auf einer universellen Zeitleiste, welches 7 Tage dauert und am 26. Oktober 2001 um 12:00 Uhr beginnt. Weiterhin definiert diese Beschreibung das Zeitleistenabbildungskonzept (*TimeLineMap*), damit die verschiedenen Zeitleisten miteinander in Relation gesetzt werden können. Dieses Konstrukt ermöglicht z.B. die Festlegung einer Verbindung zwischen einer fortlaufenden Zeitleiste eines analogen Audiosignals und deren abstrakten Zeitleiste der zugehörigen symbolischen Beschreibung des Signals. Durch diese beiden zeitbezogenen Ontologien können alle zeitspezifischen Informationen innerhalb der Musik modelliert werden [Rai08a].

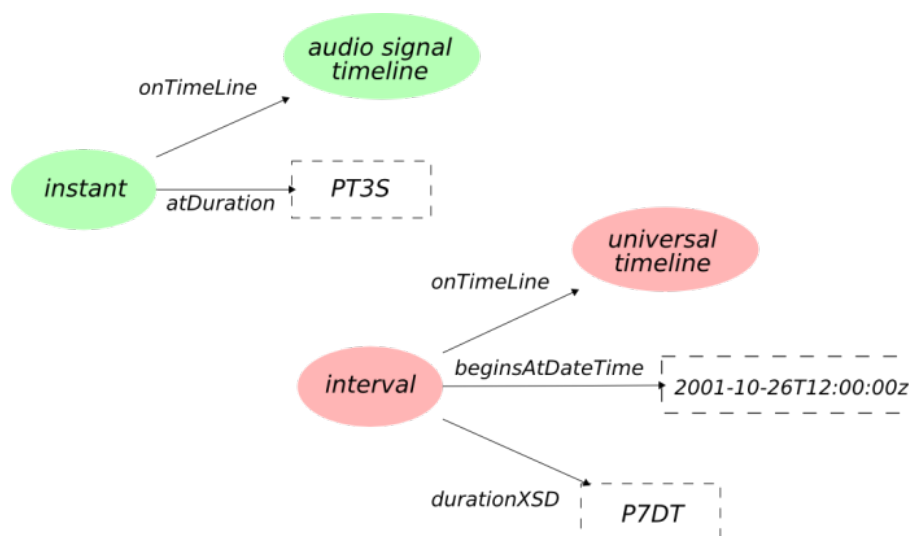


Abbildung 4.2: Beispielhafte Verwendung des Zeitleistenkonzeptes [RA07b]

Event Ontology Neben der Repräsentation von Zeitbeschreibungen mit der *Timeline Ontology* beinhaltet das *Music Ontology*-Rahmenwerk eine eigene *Event Ontology*³³.

³¹<http://purl.org/NET/c4dm/timeline.owl#>

³²<http://www.w3.org/2006/time#>

³³<http://purl.org/net/c4dm/event.owl#>

Mit ihr können Ereignisse als *Erste-Klasse-Einheiten* (*event token*[Rai08a, Gal91]) modelliert werden, d.h. ein Intervall oder eine Region beschreibt sich durch die in ihr enthaltenen Punkte (Ereignisse). Diese flexible und einfache Spezifikation (1 Konzept und 6 Eigenschaften) bietet wesentliche Vorteile gegenüber der Definition des Ereigniskonzeptes in der *ABC*-Ontologie, wo Ereignisse als Übergänge zwischen Situationen beschrieben werden. Eine Möglichkeit des Ereigniskonzeptes (*Event*) der genutzten Definition ist die Klassifikation von Regionen auf Zeitleisten, z.B. die Beschreibung des Startpunktes des 1. Kehrreims eines Liedes [Rai08a]. Weitere Vorteile können aus der Doktorarbeit von Yves Raimond[Rai08a] entnommen werden.

In Abbildung 4.3 werden die Elemente des Ereignismodells dargestellt. Als Erstes dienen die funktionalen Eigenschaften *Zeit* (*time*; mit der *OWL-Time*-Ontologie modelliert) und *Ort* (*place*; mit der *Geonames*-Ontologie³⁴ modelliert) zur Einordnung in eine zeit-räumliche Dimension. Als Zweites können einem Ereignis auch Agenten (*agents*; mit der *FOAF*-Ontologie modelliert) zugewiesen werden, z.B. Darsteller, Produzenten aber auch Computer oder Maschinen. Als Drittes besteht die Möglichkeit auch Faktoren (*factors*) zu beschreiben, d.h. beteiligte Dinge, z.B. Musikinstrumente oder Notenniederschriften. Zu guter Letzt werden die entstandenen Produkte durch die Eigenschaft *product* festgelegt, z.B. ein entstandener physischer Klang. Die Dekomposition eines Ereignisses erfolgt durch den Unterereignismechanismus (*sub_event*), womit komplexe Ereignisse in einfachere aufgesplittet werden können.

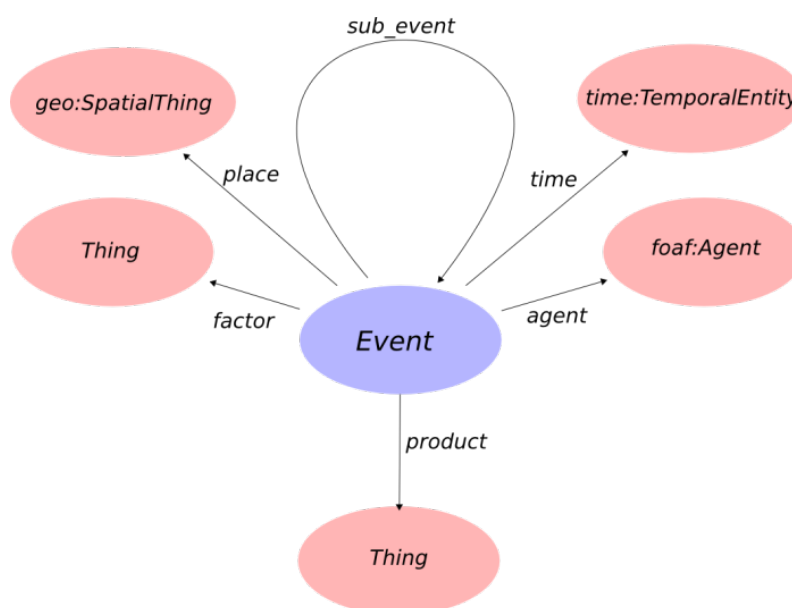


Abbildung 4.3: Das Ereignismodell der *Event Ontology* [RA07a]

Music Ontology Die eigentliche Definition der *Music Ontology*³⁵ gliedert sich in drei Ebenen (wesentliche Informationen; Entstehungsprozessinformationen; zeitlichen Erläuterungen und Ereigniszerlegung). Mit dieser Unterteilung wird die Komplexität der Beschreibungen schon durch den Standard gestaffelt und es kann damit auch der Detailgrad der Informationen durch die verwendeten Schichten dargestellt werden. Auf Basis der

³⁴siehe <http://www.w3.org/2003/01/geo/>

³⁵<http://purl.org/ontology/mo/>

vier Abstraktionsebenen der *FRBR*-Ontologie wird der Kern dieser Ontologie, die vier Grundkonzepte, definiert:

- **Musikalisches Werk** (*MusicalWork*): abstraktes, einzigartiges, musikalisches Werk bzw. Schöpfung (auch Musikwerk genannt), welche bis zum Entstehungszeitpunkt entweder noch nicht erschaffen oder noch nicht entdeckt wurde³⁶
- **Musikalische Ausdrucksform** (*MusicalExpression*): künstlerische Umsetzung eines Musikwerkes, z.B. ein physischer Klang (*Sound*), ein Audiosignal (*Signal*) oder eine Partitur (*Score*)
- **Musikalische Erscheinungsform** (*MusicalManifestation*): bündelt alle Musikstücke, welche die selben physischen und inhaltlichen Charakteristiken aufweisen und aus einer musikalischen Ausdrucksform entstanden sind, d.h. z.B. eine veröffentlichte Aufnahme eines CD-Album (i.A. *Record*), welches viele Musikstücke (*Track*) enthält, die wiederum ein Audiosignal haben
- **Musikalisches Exemplar** (*MusicalItem*): konkrete Einheit einer musikspezifischen Erscheinungsform, d.h. ein Exemplar, z.B. eine persönliche Kopie eines bestimmten veröffentlichten CD-Albums

Diese Grundkonzepte werden nun in den drei Ebenen der *Music Ontology* genutzt und spezifisch erweitert. Die erste Ebene enthält dabei wesentliche Informationen zum Beschreiben von Musikstücken ab der Veröffentlichung, d.h. grundlegende Fakten werden eindeutig auf die zugehörigen Konzepte abgebildet, z.B. Künstler (*MusicalArtist*), Band (*MusicalGroup*), Plattenfirma (*Label*) oder Musikstück [Rai08a]. Diese Schicht bietet u.a. den Einstieg zum Informieren über Hintergründe (z.B. durch Querverweise auf *MySpace*-Seiten³⁷), verfügbare Versionen des Musikstückes, ähnliche Musikstücke aber auch Verknüpfungen zu Musik-Online-Läden (z.B. *Amazon*). Dafür wird im Wesentlichen das Agentenkonzept und Eigenschaften zum Bilden von Beziehungen (z.B. Mitglied; *member*) der *FOAF*-Ontologie aufgefasst und erweitert.

In der zweiten Ebene werden alle Abläufe, die im Bezug zum Entstehungsprozess stehen, modelliert. Hiermit kann z.B. die Veröffentlichung (*Release*³⁸) einer Aufnahme (Aufnahmeereignis, *Recording*) einer Aufführung (*Performance*) eines Musikwerkes beschrieben werden [Rai08a]. Die Grundlage der Abläufe bildet demnach eine Kette von Ereignissen (z.B. Aufführung, Aufnahme, Veröffentlichung), d.h. hier wird das Ereigniskonzept der *Event Ontology* angereichert, und eine Konkretisierung der Agentenrollen vorgenommen, z.B. Komponist (*Composer*). In Abbildung 4.4 wird ein solcher Ablauf von der Entstehung bis hin zur Veröffentlichung³⁹ mit Hilfe der *Music Ontology* modelliert. Im Groben wird darauf die gerade beschriebene Konzeptverkettung schon ab der Komposition (*Composition*) dargestellt und endet in der Veröffentlichung eines Audiosignals als Tonträger (*Record*⁴⁰).

Die letzte Verfeinerung in der Beschreibungshierarchie der *Music Ontology* ist die drit-

³⁶Siehe Erläuterungen in [Rai08a, S. 50]

³⁷<http://www.myspace.com/>

³⁸z.Z. In Diskussion für die nächste Version der *Music Ontology*, siehe <http://wiki.musicontology.com/index.php/ProposalRevision14>

³⁹In der derzeitigen *Music Ontology*-Version (1.2)

⁴⁰Dieser ist in diesem Moment noch eine Art abstrakter Tonträger, welcher in Form verschiedener Audio-signalträger verfügbar gemacht werden kann, z.B. Musikdokumente (z.B. MP3), CD, *Stream*

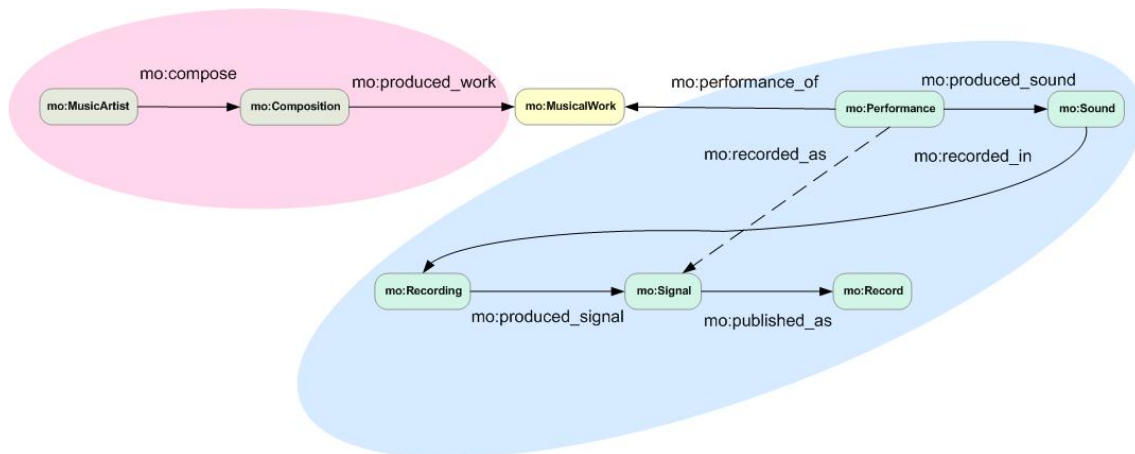


Abbildung 4.4: Der Musikentstehungsprozess mit Konzepten und Beziehungen der *Music Ontology* beschrieben [GR07]

te Ebene, welche sich mit zeitlichen Erläuterungen und der Dekomposition von Ereignissen befasst. Hiermit werden die Grundkonzepte zum Modellieren von Ergebnissen der Signalanalyse geschaffen, d.h. Musikmerkmale, wie z.B. Tempo, Lautstärke (*Loudness*) oder musikstrukturelle Beschreibungen (*MusicStructuralSegment*), wie z.B. Kehrreim und Strophe. Darüber hinaus könne aber auch detaillierte Beschreibungen von Ereignissen mit dem Unterereigniskonzept realisiert. Ein Unterereignis beschreibt z.B. die Position und die Typen von zwei Mikrofonen bei einer Aufnahme (als Oberereignis) [Rai08a]. Die wesentlichen Bestandteile dieser Ebene werden durch das Einbinden der *Event Ontology* (z.B. Unterereignis) und *Timeline Ontology* (Zeitleisten, Zeitleistenabbildungen, usw.) geschaffen. Spezifische Erweiterungen, z.B. die *Audio Feature-Ontologie*⁴¹ (z.B. *MusicSegment*) oder die *Chord-Ontologie*⁴², ergänzen diese Grundlage. Verschiedene Klassifizierungen der aufbauenden Ontologien reichern das Ereigniskonzept an und ermöglichen eine detaillierte zeitliche Beschreibung der in einem Musikstück enthaltenen Merkmale auf verschiedene Art und Weise.

Neben den schon erwähnten Erweiterungen existieren z.Z. noch die *Symbolic-Ontologie*⁴³ zur Modellierung von symbolischer Musikbeschreibung (in Form von Partituren) und die *Instrumententaxonomie* von Herman⁴⁴ als umfangreiche Ansammlung von Instrumenten, welche in einer *SKOS*⁴⁵-Hierarchie angeordnet sind und als Wurzelement das Instrumentenkonzept (*Instrument*) der *Music Ontology* wiederverwendet. Darüber hinaus existieren noch weitere Konzepte, welche durch separate und spezialisierte Ontologien erweitert werden können, z.B. das Genrekonzep (*Genre*) zur Definition einer Musikgenretaxonomie. Die *Music Ontology* bildet durch die Einbindung existierender Standardontologien und durch die ausbaufähigen Definition von Basiskonzepten eine gute Möglichkeit zur Integration in ein größeres vernetztes Wissensrepräsentationssystem, welches von einer Anwendung zur Verwaltung von persönlichen Musiksammlungen gut genutzt werden kann (vgl. [Rai08a]).

⁴¹zum Modellieren von Audiosignaleigenschaften; siehe <http://purl.org/ontology/af/>

⁴²zum Beschreiben von Akkorden; siehe <http://purl.org/ontology/chord/>

⁴³<http://purl.org/ontology/symbolic-music/>

⁴⁴<http://purl.org/ontology/mo/mit/>

⁴⁵*Simple Knowledge Organisation System*; siehe <http://www.w3.org/TR/skos-reference>

4.3 Musikmetadatendienste

Musikmetadatendienste dienen zum Einordnen, Identifizieren, Suchen und Empfehlen von Musikstücken unterschiedlichem Ursprungs (in Bezug auf den Typ des Audiosignalträgers). Diese Dienste und die dazugehörigen Datenbanken werden in Bezug auf Audio-Signal-Formate, wie z.B. *MP3* oder *Ogg Vorbis*, und für Musikstücke von z.B. CDs, DVDs und Schallplatten verwendet. Sie nutzen im Allgemeinen die zugrunde liegenden Format-/ Datenträger-gebundenen Beschreibungen, versuchen aber auch diese automatisch zu ergänzen und ggf. die Informationen durch Audiosignalanalyse anzureichern (z.B. *Last.fm*).

Eine gemeinsame Charakteristik dieser Dienste ist, dass sie jeweils eine eigene, riesige, auf die Anforderungen des jeweiligen Dienstes zugeschnittene Datenbank besitzen, welche im Laufe der Jahre stark gewachsen ist. Da die Verwendung der Beschreibungen aber oft einen kommerziellen Hintergrund hat, lassen sich manchmal nur schwer ausführliche Angaben zum Aufbau der Datenbankschemata und der verwendeten Metadatenpezifikationen machen. Dennoch sollen die vorhandenen Informationen einen Überblick über die genutzten Musikmetadaten und -merkmale bieten.

Die *Musikmetadatendienste* lassen sich in zwei grobe Bereiche gliedern, wobei ein *Musikmetadatendienst* heutzutage eine Vielzahl von Funktionalitäten bietet und die Unterteilung wohl eher in der Evolution der Versorgung der Anwender mit musikrelevanten Informationen zu begründen ist. Auf der einen Seite gibt es die *Dienste zur Katalogisierung von Audiosignalträgern* (siehe Abschnitt 4.3.1), und auf der anderen Seite die *Dienste zur Musikempfehlung und -entdeckung* (siehe Abschnitt 4.3.2).

4.3.1 Dienste zur Katalogisierung von Audiosignalträgern

Im Allgemeinen bilden die *Dienste zur Katalogisierung von Audiosignalträgern* die historische Grundlage der heutigen *Musikmetadatenstandards* (siehe 4.2), wobei sie sich zuerst nur auf die Informationsversorgung bei der Digitalisierung von herkömmlichen Audiosignalträgern, wie z.B. CDs und Schallplatten, konzentriert haben. Im Laufe ihrer Entwicklung haben sich diese Dienste aber oft so erweitert, dass sie den Benutzer mit weiteren zusätzlichen Informationen versorgen, wie z.B. Rezensionen oder auch Zuordnungen von empfundenen Stimmungen.

Compact Disc Data Base (CDDB) Die älteste Metadatenbeschreibung zum Identifizieren von CDs geht auf das Jahr 1995 zurück. Sie ist Bestandteil eines internetbasierten Datenbankdienstes (u.a. von *iTunes* genutzt) der Firma *Gracenote*⁴⁶ und mittlerweile mit in den MusicID-Dienst⁴⁷ dieses Unternehmens eingegliedert [Gra09b]. Die anfängliche CD-basierte Identifikation der Informationen für Künstler, Musikstückliste und Albumtitel wurde auf eine musikstückbasierte Identifikation erweitert (*CDDB*²), welche auf Basis von *musikalischen Fingerabdrücken* beruht (siehe Abschnitt 5.1.6). Somit kann eine genauere Einordnung der einzelnen Musikstücke vorgenommen werden und besonders die Genrekategorie besser zugeordnet werden. Diese Version liefert auch noch zusätzliche Informationen wie z.B. relevante Weblinks [Gra01]. Durch die Kommerzialisierung dieser

⁴⁶<http://www.gracenote.com/>

⁴⁷Als Teil der *Gracenote*-Produktpalette, welche sich auf das Erkennen von Musikstücken konzentriert

Technologie und der damit verbundenen Lizenzänderung haben sich mehrere Folgebeschreibungen entwickelt, welche nachfolgend beschrieben werden.

freeDB⁴⁸ Die Metadatenbeschreibung, die dieser CD-Informationsdatenbank zugrunde liegt, ist im Wesentlichen einfach nur die frei verfügbare Umsetzung des ersten *CDDB*-Formates. Da dieser Standard auf CD-basierte Identifikation beruht, ist auch der Anwendungsbereich mit Wiedergabe und Sicherung von Audio-CDs klar definiert, was sich auch in den Programmen widerspiegelt, die diesen Datenbank-Dienst unterstützen.

Mit nur 11 Genre-Kategorien ist die Auswahl zur Klassifikation der Alben sehr begrenzt, aber ein frei definierbares Feld ermöglicht es den Benutzer auch eigene Zuordnungen festzulegen [MAG06]. Da die Kennzeichnungen sich immer auf eine Veröffentlichung beziehen, kann dies tendenziell auch zu falschen Einordnungen bezüglich des Musikgenres eines einzelnen Musikstückes führen. Eine Mehrfachauswahl ist hier leider auch nicht möglich. Weiterhin unterliegen neue Einträge auch noch diversen Restriktionen (z.B. Bandnamen mit „the“ am Anfang dürfen nur ohne diesen Artikel eingetragen werden), was nicht gerade zu einer intuitiven Handhabung des Formates beiträgt, aber generell zu einer höheren Qualität der Datenbankeinträge [MAG06].

Generell ist dieses Metadatenschema von der Struktur her recht veraltet (dies trifft dann folglich auch auf die erste *CDDB*-Version zu). Durch den Einstieg der Firma *MAGIX*⁴⁹ in das Projekt können sich in Zukunft eventuell noch Verbesserungen und Weiterentwicklungen ergeben.

MusicBrainz XML Metadata Format(MMD) Dies ist der Metadatenstandard von der frei verfügbaren, von Benutzer gepflegte Musikmetadatenbank *MusicBrainz*, welche auch ein Nachfolgeprojekt von *CDDB* ist. Neben den üblichen Beschreibungen für ein Aufnahme, können auch persönliche Schlagwörter vergeben, Bewertungen durchgeführt und Hintergrundinformationen über die Künstler, Plattenfirmen, Alben oder Musikstücke hinzugefügt werden. Des Weiteren ist es möglich, diese Informationen mit Querverweisen zu verbinden [mis08d].

Der Vorläufer dieser Formates war eine *RDF*-Beschreibung, welche aber durch ihre Komplexität auf nur wenig Benutzerakzeptanz stieß [May07b]. Der aktuelle Standard soll einfacher zu verarbeiten und erweiterbar sein. Er basiert auf einer *Relax-NG-Schema*-Beschreibung⁵⁰. Ein von diversen *Client*-Anwendungen genutzte Schnittstelle zur *MusicBrainz*-Datenbank, ist der *REST-Webservice* [mis08c], wobei die dadurch bereitgestellte Daten auf diesem Format basieren. Zusätzlich können auch Anwendungen, welche den *freeDB*-Datenbankdienst nutzen den *MusicBrainz*-Datenbankdienst als Informationsquelle integrieren [May07a].

Ein gern genutzte Eigenschaft dieses *Musikmetadatenstandards* ist nicht zuletzt der *übertragbare einzigartige Identifikator* (engl. *Portable Unique Identifier*; *PUID*), welcher eine 128-bit-lange Repräsentation eines *musikalischen Fingerabdruckes* (siehe Abschnitt 5.1.6) ist. Mit diesen Kennzeichner können insbesondere mit wenig *redaktionellen Metadaten* versehene Musikstücke anhand ihrer Audiosignalbeschreibung analysiert und ggf. zugeordnet werden (siehe [mis09d]).

⁴⁸<http://www.freedb.org/>

⁴⁹<http://www.magix.com/>

⁵⁰Eine XML-Schemabeschreibung, siehe <http://www.relaxng.org/>

Discogs⁵¹ Speziell zum Auflisten kompletter Schallplatten-/ Musik-CD-Sammlungen bzw. (Neu-)Veröffentlichungen wurde diese von Benutzern gepflegte Datenbank mit der dazugehörigen Metadatenbeschreibung entworfen, ist aber auch auf digitale Musiksammlungen anwendbar. Anfänglich wurde mit diesem Dienst nur Musik aus dem elektronischen Musikbereich abdeckt (was auch an der großen Anzahl von Musikstildefinitionen in diesem Bereich erkennbar ist), mittlerweile wurden die Genred Definitionen aber auf 15 Grundkategorien erweitert, welche den gesamten Bereich von Musikgenreklassifikationen abdecken⁵². Somit ist dieses Format ein durch die Entwicklung dieser Anwendung (Datenbank) gewachsener Standard.

Durch den besonderen Anwendungsbereich gibt es zusätzliche *Metadatenfelder* neben den Gebräuchlichen für Künstler, Plattenfirma und Titel der Veröffentlichung und deren enthaltenen Musikstücktitel. Es wird der Katalognummer, dem Veröffentlichungsdatum, dem Veröffentlichungsland und dem Format des Musikdatenträgers eine hohe Bedeutung zugewiesen. Die Genre- und Stilzuweisungen⁵³ können mehrfach erfolgen, sind aber beschränkt auf die Veröffentlichung und nicht für einzelne Musikstücke. Weiterhin können auch die an der Produktion/ den Musikstücken beteiligten Künstler inkl. deren Rolle (z.B. Saxofonist; die Rollenbezeichnungen sind aber nicht fest definiert) und Coverbilder zu den Musikdatenträgerbeschreibungen hinzugefügt werden [Dis09]. Rezensionen, Empfehlungen und Bewertungen runden das Informationsangebot zu einer Veröffentlichung ab. Letztendlich können auch noch Informationen über den Künstler (z.B. verwendete Alias-Namen) oder der Plattenfirma (z.B. Ableger des Musiklabels) hinterlegt werden. Über einen *XML*-basierten *REST-Webservice*⁵⁴ kann der *Discogs*-Datenbankdienst auch von anderen Anwendung genutzt werden. So werden z.B. die Informationen teilweise mit in den *MusicBrainz*-Datenbankdienst (siehe 4.3.1) eingespeist bzw. durch diesen miteinander verknüpft, oder in Musikwiedergabe- und Verwaltungsanwendungen, wie z.B. *Songbird*, repräsentiert.

All Music Guide Eine ähnlich historisch-gewachsene Metadatenbeschreibung (Ursprung im Jahr 1991) wie *CDDB*, ist die von *All Music Guide (AMG)*. Der wesentliche Unterschied zu den vorherigen Beschreibungen ist dabei aber, dass hier der Inhalt von professionellen Kommentatoren erzeugt wird und nicht von einer Menge von Benutzern⁵⁵. Das Format ist in vier wesentliche Kategorien unterteilt, welche im Folgenden kurz beschrieben werden [Mac09a]:

- **Metadaten**⁵⁶: Fakten über das Album oder den Künstler (inkl. Musikstücktitel, Musikstückanzahl, Genre, Plattenfirma)
- **Beschreibender Inhalt**: Detaillierte Informationen zum Beschreiben des Albums oder des Künstlers (inkl. Musikstil, Stimmungen, Instrumenten, Herkunftsland)

⁵¹<http://www.discogs.com/>

⁵²siehe http://wiki.discogs.com/index.php/Style_Guide-Related_Styles_Lists

⁵³Stil hier im Sinne von Unterkategorie eines Genres

⁵⁴<http://www.discogs.com/help/api>

⁵⁵Im Fall von *CDDB* überwachen mittlerweile auch professionelle Kommentatoren die Einträge mit [Gra09a]

⁵⁶Hier als Kategorie zu verstehen

- **Relationaler Inhalt:** Informationen zum Herstellen von Querverweisen zwischen den Künstlern und ihrer Musik (z.B. *die den Künstler beeinflussenden Künstler* oder *ähnliche Künstler*)
- **Redaktioneller Inhalt:** Hintergrundinformationen wie z.B. Künstlerbiografien, Musikstück- oder Albenrezensionen, oder Musikstilbeschreibungen

Durch seine lange Evolution und recht breit angelegte Struktur ist dieses Metadatenchema zu einer Art De-Facto-Standard geworden. Leider ließ sich durch den kommerziellen Hintergrund dieses Systems keine genau Beschreibung der genutzten Spezifikation finden. Eine wesentliche Unterteilung ist durch die Abgrenzung zwischen populärer und klassischer Musik zu erkennen, welche als separate Datenbanken erhältlich sind. Darüber hinaus werden den Musikalben oder -zusammenstellung empfundene Stimmungen und passende Ambiente (Themen) zugeordnet. Eine aktuelle Statistik des Umfangs der Informationen über populäre Musik zeigt deren hohe Abdeckung (siehe Tab. 4.1). Interessant ist hierbei die große Anzahl an Genres bzw. Musikstilen, Stimmungen und Themen, da diese *Metadatenfelder* u.a. die Basis für Technologien zur Wiedergabelistengenerierung bilden aber auch eine andere Sicht auf die persönliche Musiksammlung ermöglichen.

Alben	1.464.083	Genres und Musikstile	926
Musikstücke	14.274.193	Themen	86
Albumrezensionen	326.790	Stimmungen	184
Biographien	87.941	Instrumente	5.044
Personen (alle Namen)	1.082.761	Querverweise	10.632.067
Verfasser	247.124	Coverbilder	948.120
Albumquellennachweise	9.389.914	Künstlerfotos	68.232

Tabelle 4.1: Abdeckungsstatistik der Datenbank für populäre Musik von AMG [Mac09c]

Die Lizenzen für die Metadaten werden je nach Umfang gestaffelt und reichen von grundlegenden Produktdaten (u.a. Künstler, Titel, Genre) bis zu Komplettpaketen, die alle beschriebenen Kategorien enthalten, u.a. Ähnlichkeiten (für Künstler, Musikstil, Instrumente usw.) und Hintergrundinformationen (siehe [Mac09b]). Besonders *Online*-Musikläden nutzen die Datenbank des *All Music Guide* gern (z.B. *Amazon* oder *Napster*). Dafür bietet die Firma *Macrovision* eine spezielle Software (*LASSO*) zur Analyse von Audiomaterial (Musikdokumente, CDs und DVDs), welcher u.a. auch von Elektronikgeräteproduzenten genutzt werden kann [Mac09d].

4.3.2 Dienste zur Musikempfehlung und -entdeckung

Diese Kategorie von Dienste versorgt den Benutzer in der Hinsicht mit Musikmerkmalen- und Metadaten, dass dieser bestimmte Angaben seines Geschmacks auf unterschiedliche Weise machen kann, um darauf Musikempfehlungen zu bekommen, wobei er auch ggf. neue Musik entdecken kann. Dieser Ansatz bildet auch oft die Grundlagen für Anwendungen zur Wiedergabelistengenerierung (siehe Abschnitt 2.5). Die Informationsgrundlagen, welche für die Musikempfehlung und -entdeckung genutzt werden, können dabei unterschiedliche Ansätze verfolgen und charakterisieren sich dabei meistens durch Bewertungen und Einordnungen von Benutzern und/oder professionellen Kommentatoren.

Ihre Einsatzgebiete sind aber recht verschieden, z.B. Kaufempfehlungen, angepasstes Internetradio oder die Zusammenstellung von Musik für ein bestimmtes Ambiente.

Last.fm Der wohl bekannteste Online-Dienst, welcher angepasste (Internet-)Radio-stationen⁵⁷ ermöglicht, ist *Last.fm*. Die für die Musikempfehlung- und entdeckung verwendete Ähnlichkeitsanalyse ist das *gemeinschaftliche Filtern* (siehe Abschnitt 2.4), welches auf den, durch Kennzeichnungen der Benutzer (März 2009 ca. 30 Mill. Benutzer [Jon09]), beschriebenen Musikstücken, -alben und -künstlern, beruht und deshalb mit den Nachteilen dieses Ansatz behaftet ist. Eine Mehrfachauswahl der Startattribute (Musikkünstler, Schlagworte) zur Erstellung einer personalisierten Radiostation ist dabei möglich. Neben einer sehr hohen Benutzeranzahl zeichnet sich die Beliebtheit dieses Musikmetadaten dienstes auch durch die hohe Loyalität der Anwender aus, wobei nach Angaben von *Last.fm* 75% der Benutzer innerhalb von 24 Stunden den Dienst wieder benutzen [Hym08].

Dieser *Musikmetadaten dienst* versorgt die Anwender nicht nur mit den genannten Basisdaten zur Musikempfehlung von über 70 Mill. Musikstücken, 12 Mill. Musikkünstlern und 20 Mill. (unstrukturierten) Schlagwörtern, sondern mit noch weiterem Hintergrundwissen. So gibt es ca. mehr als 200.000 *Wiki*-Seiten über Musikkünstler und über 500.000 gelistete Veranstaltungen (weltweit), welche durch die Anwendergemeinschaft, den Plattenfirmen oder den Künstlern selbst, oder professionellen Veranstaltungspartnern eingetragen werden (siehe [Hym08]).

Die Musikstückinformationen (Künstlernamen, Ablumtitel, Musikstücktitel, usw.) werden dabei oft direkt von den Benutzern während des Hörens gesammelt, in dem diese Metadaten automatisch an den *Last.fm*-Dienst gesendet werden. Darüber hinaus kann mit der *Last.fm-Client*-Software ein *musikalischer Fingerabdruck* (siehe Abschnitt 5.1.6) erstellt werden, aus welchem eine einzigartiger Identifikator pro Musikstück erstellt wird und dieser somit auch zur Korrektur von eventuell falsch geschriebenen Bezeichnungen genutzt werden kann (siehe [Las09]).

Eine richtig vollständige Qualitätskontrolle ist diese Technik aber noch nicht und es gelangen schnell falsche Bezeichnungen wie z.B. „NULL“ in die persönliche Bibliothek. Semantische Zusammenhänge in den *Metadatenfeldern* und Schlagwörtern können z.Z. leider noch nicht erkannt werden⁵⁸, z.B. mehrere Künstlernamen (Kooperationen usw.) im dafür vorgesehenen *Metadatenfeld* oder unterschiedliche Schreibweisen bei den Kennzeichnern (z.B. „hip hop“, „hiphop“, „hip-hop“).

Letztendlich kann der gesamte Informationsaustausch z.B. über den *REST-Webservice* der *Last.fm-API*⁵⁹ abgewickelt werden, so dass zahlreiche Anwendungen das vorhandene Wissen in unterschiedlichen Formen repräsentieren (z.B. *Songbird's mashTape*⁶⁰) und auch so Informationen an den *Last.fm*-Dienst übermittelt werden können.

Pandora *Pandora*⁶¹ ist eine Internetradiofirma, welche, ähnlich wie *Last.fm*, angepasste *Radiostationen* erstellt, die auf einem oder mehreren Startattributen basieren, wobei diese entweder Musikstücke oder Künstler sein können. Der wesentliche Unterschied

⁵⁷Als eine Form einer automatisch generierten Wiedergabeliste, siehe Abschnitt 2.5

⁵⁸Die *Wiki*-Seiten können aber mit semantischen Kennzeichnern beschrieben werden, z.B. *Künstler*

⁵⁹<http://www.last.fm/api/>

⁶⁰<http://addons.songbirdnest.com/addon/73>

⁶¹<http://www.pandora.com/>

zwischen dem kollaborativem Ansatz von *Last.fm*, ist der, dass hierbei jedes Musikstück von Musikexperten (und auch nur von diesen) mit einer passenden Charakterisierung, auf einer definierten Menge von ca. 400 Eigenschaften, beschrieben wird. (siehe [Cli07]). Jede Eigenschaft wird auf einer 10er-Skala bewertet und es sind u.a. die Instrumente, deren Klang (z.B.), Rhythmus und Harmonie, der Inhalt des Liedtextes (z.B. ein trauriger Liedtext) und die zugehörige Stimme, die grundlegende Stimmung und der passende Musikstil (z.B. *Einflüsse von Funk*) zu beurteilen. Die Charakteristiken sind dabei recht anschaulich gewählt bzw. kann deren Detailgrad variieren, z.B. eine einfache Beschreibung „männliche Stimme“ oder eine komplexe Beschreibung „aggressive männliche Stimme“ (siehe [mis09f]). Neben der Auswahl der Startattribute kann der Benutzer durch Anweisungen, wie z.B. „Daumen hoch“ oder „Daumen runter“, versuchen die *Radiostation* bei der Ähnlichkeitsanalyse und Musikstückauswahl zu beeinflussen (siehe [Cli07]). Obwohl die Firma schon seit 1999 versucht ihr Expertenbewertungskonzept für Musikempfehlungen zu vermarkten und gerade durch die 2005 eingeführte Plattform für die personalisierten Radiostationen einen enormen Anwenderzuwachs bekommen hat (Ende 2008 ca. 20 Mill. registrierte Benutzer[Sch08]), muss sie trotzdem noch mit dem Problemen kämpfen die dieser Ansatz beinhaltet. Auf der einen Seite ist es schwer mit einer geringen Anzahl von Musikexperten möglichst viele Musikstile mit möglichst vielen Musikstücken abzudecken. Auf der der anderen Seite muss auch der Grad der Objektivität, welchen es zu erreichen gilt, in Frage gestellt werden, durch die letztendlich subjektive Bewertung der Musikexperten. Letztendlich kann auch nicht genau bestimmt werden, wie flexibel die Menge der verwendeten Kennzeichner auf die aktuelle Musikentwicklung angepasst wird⁶².

Rockanango Projekt Als gemeinsame Entwicklung der Katholike Universiteit Leuven und des Musikinhaltsanbieter *Aristo Music*⁶³ entstand im Rahmen des *Rockanango*-Projektes [CGD06] ein *Musikmetadaten*schema mit besonderem Augenmerk zur Musikempfehlung und Wiedergabelistenerzeugung im *HORECA*⁶⁴-Umfeld. Das wesentliche Gestaltungsziel war dabei die richtige Musik für eine bestimmte Zeit und einen bestimmten Ort anzubieten, d.h. Musik die auf eine gewisse Situation und eine gewünschte Atmosphäre passt [CGD06]. Diese Idee wird hierbei als *Musikkontext* bezeichnet und setzt sich aus einer Menge von gewichteten *Untermusikkontexten* zusammen, d.h. es können verschiedene Grunddefinitionen miteinander vermischt werden um eine vielfältige Wiedergabeliste entstehen zu lassen.

Durch die Wertebelegung der Musikparameter innerhalb eines festen Wertebereiches können Musikstücke den anpassbaren *Musikkontexten* zugeordnet werden, wenn sie mit deren Kriterien übereinstimmen. Der Vorteil in dieser Technik liegt darin, dass der Anwender durch grobe Beschreibungen seiner Vorstellung der gewünschten Musik, d.h. der Auswahl, Erstellung oder Anpassung eines *Musikkontextes*, diese aus der Auswahl einer größeren Musiksammlung⁶⁵ präsentiert bekommt. Ein Tanzabend kann beispielsweise wie folgt beschrieben werden [CGD06]:

- **Genre:** Dance

⁶²Es existieren aber unvollständige Listen der Eigenschaften (siehe [mis09f, mis09g])

⁶³<http://www.aristomusic.com/>

⁶⁴**HO**otels, **RE**staurants, **CA**fes usw., d.h. Musikwiedergabe in einem weitestgehend öffentlichen Raum

⁶⁵2007 ca. 30.000 Musikstücke [GCD07]

- **Stimmung:** fröhlich, überschwänglich, außer sich, stimulierend
- **Bekanntheitsgrad:** mittelmäßig bis sehr hoch
- **Partyfaktor:** mittelmäßig bis sehr hoch
- **Lautheit:** mittelmäßig bis hoch

Neben diesen Musikparameter haben u.a. noch die folgenden *Metadatenfelder* eine einflussreiche Bedeutung zur Generierung von *Musikkontexten*: Tempo, Rhythmusstil, *Tanzbarkeit*, Aktualität und geografische Region. Problemfälle bei der Modellierung sind die Elemente Untergenre und allgemeiner Bekanntheitsgrad. Auf der einen Seite können die Musikstile als gewichtete Mehrfachbelegung dem Musikparameter Genre zugeordnet werden. Auf der anderen Seite sollte vielleicht der allgemeine Bekanntheitsgrad durch eine zielgruppenspezifische Unterteilung ersetzt werden, damit diese nicht nur den Durchschnittsanwender repräsentiert und dadurch eine größere Akzeptanz erreicht wird⁶⁶.

Eine Erweiterung dieses Musikmetadatenschemas fand hinsichtlich der Unterstützung von automatischer Metadatengenerierung, als Hilfsmittel für die manuell kommentierende Gruppe von Musikexperten⁶⁷, statt (siehe [GCD07]). Einerseits ergeben sich durch inhaltsbasierte Methoden der Schlagwortextraktion⁶⁸ aus den Liedtexten der Musikstücke die Werte für das Metadatenfeld Musikstückthema. Dies eröffnet dem Anwender die Möglichkeit, nach Musikstücken durch die Eingabe von Liedtextauszügen zu suchen, welche zu mindest phonetisch eine gewisse Übereinstimmung aufweisen müssen. Andererseits sollen die Parameter Grundtonart und *Modus* des Musikstückes (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“), welche sich aus der Akkordextraktion ergeben, helfen, die stimmungsrelevanten Parameter vorauszuberechnen. Datenbankauswertungstechniken (siehe Abschnitt 5.2.3), wie z.B. der *Apriori*-Algorithmus, helfen dabei, verwertbare Assoziationsregeln zu erstellen, z.B. aus einer geringen *Tanzbarkeit* lässt sich auch ein geringer Partyfaktor ableiten [GCD07].

Letztendlich wird durch die Einführung von Kategorien mit einem beschränkten Wertebereich die flache Modellierung aufgelöst, lässt aber gleichzeitig auch größere Angaben zu. Ein Anwender kann durch Angaben wie „diese Latinvariante“ den gewünschten Musikstil bekommen ohne dabei die genauen Bezeichnungen wie z.B. Cumbia⁶⁹ oder Rumba zu kennen [GCD07]. Zusammenfassend können diese, eigentlich für einen anderen Anwendungsbereich entworfenen *Metadatenfelderdefinitionen* in teilweise modifizierter Form gut in die Modellierung eines Metadatenschemas für persönliche Musiksammlungen übernommen werden.

Amazon Der Online-Versandhandel *Amazon* bietet neben seinem Hauptgeschäftszweig, dem Buchhandel, auch eine Plattform zum Handel mit diversen anderen Produktlinien – insbesondere auch analoge und digitale Audiosignalträger jeglicher Art. Für Kaufempfehlungen nutzt dieser Dienst, ähnlich wie *Last.fm*, das *gemeinschaftliche Filtern* (siehe Abschnitt 2.4) und bietet z.B. Produkte an die andere Kunden in diesem Zusammenhang

⁶⁶Ohne Betrachtung der schwierigeren Ermittlung eines zielgruppenspezifischen Bekanntheitsgrades

⁶⁷Der Musikexperte als qualitätserhaltende Kontrollinstanz

⁶⁸z.B. auf Techniken der Computerlinguistik beruhend

⁶⁹Ein Paartanz aus Kolumbien

gekauft haben. Er bietet die gebräuchlichen Musikmetadaten über die Musikzusammenstellungen (u.a. Veröffentlichungsdatum, Coverbilder und Musikgenre), und darüber hinaus Autoren- und Kundenrezensionen, Kennzeichnung mit Schlagwörtern und die Popularität der Audiosignaldatenträger [CGVD08].

Generell sind die verfügbaren *redaktionellen Metadaten* hochwertig aber in ihrer Reichweite durch den Handel etwas beschränkt, obwohl sich die Plattform auch für den privaten Bereich und Kleinvertriebe geöffnet hat (siehe [Ama09a]). Eine umfangreiche Taxonomie mit Musikgenres und -stilen hilft dem Benutzer bei der Suche und Einordnung von Musik. Die Konstruktion der Taxonomie scheint aber etwas misslungen zu sein, da z.B. gewisse Bezeichnungen in der Unterkategorie *Vinyl Records* erneut auftauchen und eine Musikgenretaxonomie generell unabhängig vom gewählten Audiosignalträger sein sollte.

Die *Product Advertising API*⁷⁰ bietet Programmierern über diverse *Webservice*-Schnittstellen (*REST*, *SOAP*) den Zugang zu den Produktdaten der *Amazon*-Plattform mit all ihren Entdeckungs- und Empfehlungsmöglichkeiten, ist aber gleichzeitig auch auf den kommerziellen Handel dieser ausgerichtet. Deshalb ist der Einsatz dieses *Musikmetadaten*-dienstes für eine Anwendung, bei der die kommerzielle Orientierung nicht unbedingt im Vordergrund steht, nicht gut geeignet (sofern dies überhaupt mit den Lizenzbedingungen vereinbar ist, siehe [Ama09b]).

4.4 Fazit

Die in diesem Kapitel vorgestellten *Musikmetadatenformate* und *-dienste* geben einen Überblick über oft genutzte, aktuelle und zukunftsweisende Datenmodelle und Informationsquellen für Musikmerkmale und -metadaten. Wie die wesentlichen Anwendungsfälle im privaten Umfeld in Kapitel 2 gezeigt haben, muss eine Anwendung zur Verwaltung von persönlichen Musiksammlungen, eine Vielzahl von Informationen über die Inhalte und Zusammenhänge der in der Kollektion enthaltenen Musikstücke, -zusammenstellungen und Künstler enthalten, damit der Benutzer einen vielfältigen Zugang zu dieser ermöglicht bekommt. Wobei die in Kapitel 3 beschriebenen charakteristischen Eigenschaften und Kategorisierungen eine Basis bei der Musikinformationsversorgung bilden sollen und mit den weitestgehend kontextuellen Informationen, die in Musikedokumenten enthalten sind und über *Musikmetadiensten* erhältlichen Beschreibungen, eine Einheit bilden.

Empfehlungen für formatgebundene Spezifikationen Die vorgestellten *formatgebundenen Spezifikationen* zeigten entweder deutlich ihre Grenzen in der Ausdruckskraft aus (z.B. *ID3*) oder können sich durch einen Mangel an Standardisierung nicht durchsetzen (z.B. *Vorbis Comments*). Neben der eventuell guten Unterstützung von möglichst vielen *redaktionellen Metadaten* (sofern denn viele der zusätzlich empfohlene *Metadatenfelder* genutzt werden, z.B. auch *Entstehungsjahr*, *Entstehungsland*, *beteiligte Künstler*, *Veröffentlichungsdatum*) in diesen Spezifikationen, ist eine Einbindung von *akustischen Metadaten* (wie z.B. *Lautheitsangaben*, *BPM*, *Instrumentierung*) empfehlenswert. Eingebettete Coverbilder sind zur visuellen Repräsentation von Musikedokumenten von hoher Bedeutung und ermöglichen eine einfachere Identifikation dieser.

⁷⁰<https://affiliate-program.amazon.com/gp/advertising/api/detail/main.html>

Eine Mehrfachbelegung von bestimmten *Metadatenfeldern* wie z.B. Musikgenre oder Stimmung sollte möglich sein, wobei diese eventuell noch jeweils eine bestimmte Gewichtung (siehe z.B. der von *Pandora*) bekommen könnte. Weiterhin muss eine Separierung und genaue Benennung (z.B. Saxophonist) der an einem Musikstück beteiligten Künstler möglich sein, damit semantische Zusammenhänge erhalten bleiben. Es empfiehlt sich daher auch bei den *formatgebundenen Spezifikationen* bzw. bei den *Musikmetadaten*, welche direkt im Musikdokument enthalten sind, der Einsatz eines ontologiebasierte Metadatenformates, wie z.B. die *Music Ontology*, da damit z.B. dereferenzierbare *URIs* anstatt nur einfacher textueller Beschreibungen genutzt werden können. So können problemlos weitere Hintergrundinformationen bezogen werden.

Darüber hinaus könnten die *akustischen Metadaten* ggf. durch detaillierte Merkmalsvektoren und -matrizen aus der Audiosignalextraktion- und anreicherung (siehe Abschnitt 5.1) dem eigentlichen Audiosignal, welches in komprimierter oder unkomprimierter Form vorliegt, ergänzt werden, um eine spezifische Inhaltsauswertung weiterhin zu ermöglichen (wobei das Audiosignal in persönlichen Musiksammlungen oft komprimiert ist). Zur Steigerung der Qualität der weiteren Inhaltsauswertungen empfiehlt es sich eventuell, brauchbare Merkmalsvektoren und -matrizen des Originalaudiosignales zu verwenden, welche bei der Erstellung einer komprimierten Version dem Musikdokument beigelegt werden können und ggf. separat neben denen des komprimierten Audiosignals eingefügt werden. Diese Merkmalsvektoren und -matrizen sollten letztendlich mit den hochwertigen, abstrakteren Eigenschaften, welche durch die Ontologie beschrieben sind, verknüpft werden können, so dass diese dadurch über ein Element eines *semantischen Graphens* zugänglich sind.

Letztendlich sollten die *Musikmetadaten* (und -merkmalsbeschreibungen), welche direkt in den Musikdokumenten enthalten, eher objektiv sein, damit diese Informationen auch gut wieder benutzt werden können. Eine subjektive Anpassung von Eigenschaften, sollte separat in einer lokalen *Musikwissenbasis* (siehe Abschnitt 5.2.4) vorgenommen werden.

Empfehlungen für formatunabhängige Spezifikationen Die *formatunabhängigen Spezifikationen* zur Beschreibung von Musik, sollten letztendlich alle *Musikmetadatentypen* abdecken und in der lokalen *Musikwissenbasis*, welche die Grundlage zur Verwaltung einer persönlichen Musiksammlung bilden soll, zum Einsatz kommen. Wo bei den *MPEG*-Standards noch mindestens zwei separate Standards benötigt werden um die Anwendungsfälle im privaten Umfeld zu modellieren, wird durch die Einführung von ontologiebasierten Metadatenschemata, z.B. auf Basis der *Music Ontology*, eine in allen Bereichen erweiterbare Grundlage gelegt, in der auch die semantischen Zusammenhänge erhalten bleiben. Weiterhin kann vorhandenes Wissen, welches mit Hilfe von Ontologien beschrieben wurde, gut wiederverwendet werden und durch den Einsatz diverser De-Facto-Ontologien, wie z.B. *FOAF*, einfach integriert werden.

Dies unterstützt den Prozess der Merkmalsextraktion (siehe nächstes Kapitel), welcher neben der Rückgewinnung von inhaltlichen und redaktionellen Musikbeschreibungen direkt aus dem Musikdokument, diese Charakteristiken auch noch durch weitere Informationsquellen, i.A. verschiedene *Musikmetadaten Dienste*, mit Hintergrundwissen anreichert – insbesondere auch mit *kulturellen Metadaten* (vgl. [Rai08a]). Wenn nun die in einem Musikdokument vorhandenen Metadaten durch einen *semantischen Graphen* teilweise oder vollständig beschrieben werden, können diese Merkmale recht einfach in eine ontologiebasierte Wissensbasis übernommen werden und z.B. vorhandene *URIs* zur weiteren

Informationsanreicherung dereferenziert werden. Nicht vorhandene oder neue berechnete bzw. angefragte⁷¹ Merkmalsvektoren und/oder -matrizen können so auch in der lokalen *Musikwissensbasis* gehalten werden und ggf. auch wieder zurück in das Musikedokument geschrieben werden. Dabei sollte ein *Metadatenfeld* vorhanden sein, welches einen eindeutigen Identifikation enthält, der aus einem *musikalischem Fingerabdruck* (siehe Abschnitt 5.1.6) berechnet wurden.

Auch die von den Anwendungsfällen benötigten Benutzerprofile für eine bessere, individuellere Bearbeitung der Aufgaben, können so einfach in die *Musikwissensbasis* integriert werden, um Beschreibungen wie z.B. *Musikkontext* (siehe *Rockanango*-Projekt) einfach zu ermöglichen. Wichtig ist in diesem Zusammenhang auch die Erstellung einer Hörerhistorie, d.h. insbesondere das Vorhandensein von *Wiedergabe-* und *Überspringzählern*, welche mit in die Berechnung von z.B. Musikempfehlung oder Wiedergabelistengenerierung einfließen können. Dies heißt nicht zuletzt, dass eine *formatunabhängige Spezifikation* zur Verwaltung von Musikinformationen, neben (weiteren) *objektiven Metadaten* insbesondere auch *subjektive Metadaten* modellieren können muss, welche lokal, aber immer noch unabhängig von einer spezifischen Endbenutzeranwendung (z.B. eine einfache Anwendung zur Musikwiedergabe), gehalten werden sollten.

Der Vorteil in der Anwendung von Ontologien liegt auch darin, dass über die vorhandene Wissensbasis recht effizient entschieden werden kann, um z.B. ungenaue Suchanfragen zu bearbeiten, und die jeweils unterschiedliche Menge an Informationen für die Anwendungsfälle gut erreicht werden kann. Somit kann dem Benutzer und Hörer einer persönlichen Musiksammlung eine Vielfalt an Musikmerkmalen und -metadaten repräsentiert werden, welche zu einem zufriedenstellenden Hörerlebnis führen können und verschiedene Benutzertypen abdecken, z.B.

- Hörer, die nur nebenbei hören (benötigen nur grundlegende Musikstückinformationen),
- Hörer die gezielt Musik hören und wahrnehmen (benötigen schon detaillierte Musikinformationen, z.B. Beschreibungen von Musikmerkmalen),
- Hörer, die darüber hinaus auch noch an weiteren Hintergrundinformationen (z.B. Liedtext) interessiert sind oder
- Hörer, welche ihre Musiksammlung neu und/oder aus einem Blickwinkel erleben möchten (benötigen die Informationen, welche sie bisher nicht bekommen konnten).

Eine einheitliche Unterteilung an vorgeschriebenen und optionalen Musikmerkmalen und -beschreibungen kann deshalb nicht gemacht werden, es können eventuell nur Vorlagen für Benutzertypen erstellt werden, welche gewissen *Metadatenfeldergruppierungen* enthalten aber immer noch anpassbar bleiben müssen. Mit diesen Vorlagen kann eine grundlegende Merkmalsextraktion vorgenommen werden, welche bei der Anpassung oder dem Wechsel des Benutzertypen aufgefrischt wird.

⁷¹Wenn die inhaltlichen Beschreibungen schon in einer lokalen oder über das Internet zugänglichen Datenbank verfügbar sind

5 Merkmalsextraktion

Die Merkmale von Musiksignalen und die Metadaten von Musikdokumenten (bzw. Audiosignalträgern) wurden bereits in den Kapiteln 3 und 4 vorgestellt. Es soll hier nun auf verschiedene Techniken eingegangen werden, mit welchen die Merkmale aus dem Signal extrahiert werden (siehe Unterkapitel 5.1). Auf der anderen Seite sollen aber auch Methoden vorgestellt werden, mit denen Metadaten der zugrunde liegenden Formate in ein gemeinsames Schema übergeführt und zusätzlich auch angereichert werden können (siehe Unterkapitel 5.2). Eine effiziente Audiosignalextraktion und Metadatenanalyse bilden die Grundlage für eine persönliche *Musikwissenbasis*, welche die Anforderungen der Anwendungsfälle auf Musiksammlungen im privaten Umfeld erfüllt (siehe Kapitel 2).

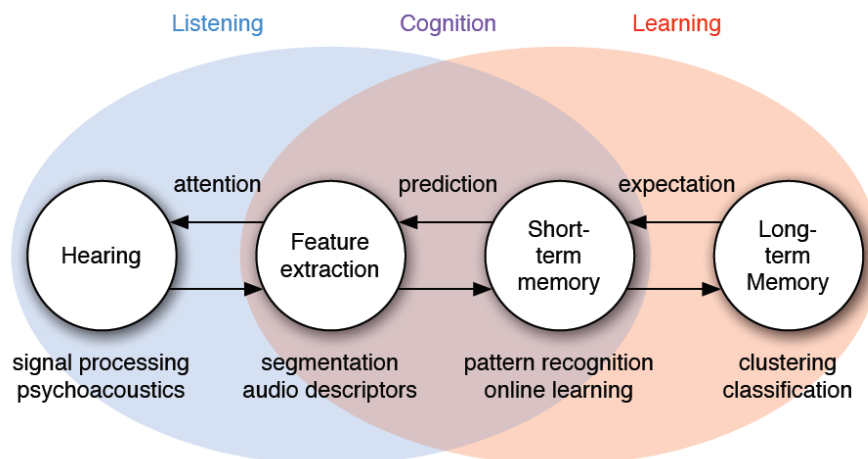


Abbildung 5.1: Musiksignalanalyse-Rahmenwerk von Jehan [Jeh05]

Im Rahmen von Jehan's Doktorarbeit [Jeh05] entstand ein Entwurf eines *Musiksignal-analyse-Rahmenwerkes* (siehe Abb. 5.1), welches die einzelnen Bereiche der Audiosignalextraktion und Metadatenanalyse abbildet. Es gliedert sich in die vier Blöcke „Gehör“ („Hearing“), „Merkmalsextraktion“ („Feature Extraction“), „Kurzzeitgedächtnis“ („Short-term memory“) und „Langzeitgedächtnis“ („Long-term memory“) auf. Dabei können die ersten drei Blöcke dem Bereich „Hören“ („Listening“; siehe Abschnitt 5.1) und die letzten drei Blöcke dem Bereich „Lernen“ („Learning“; siehe Abschnitt 5.2) zugeordnet werden. Die Schnittmenge dieser beiden Bereiche bildet dabei die *Musikkognition* („Cognition“), bei welcher neben dem eigentlich Audiosignal aber auch noch andere Bereiche, wie z.B. kulturelle Eigenschaften oder Emotionen Einfluss haben (siehe [Jeh05]).

Die Menge der zu verarbeitenden Daten, d.h. auch die Dimensionen der Vektoren, wird dabei durch jeden Block reduziert. In der ersten Ebene wird in diesem *Analyse-Rahmenwerk* mittels wahrnehmungsbasierter Frequenzfilter ein *Hörspektrogramm* (ein Spektrogramm auf Basis der Bark-Skala; siehe Abschnitt 3.1.1.2) erstellt, d.h. es findet eine Frequenzanalyse statt. Danach werden perzeptuelle Eigenschaften (siehe Abschnitt 3.1.2.1)

zu einer Art symbolischer Musikrepräsentation („musical DNA“ [Jeh05] bzw. *musikalischer Fingerabdruck*¹) zusammengefasst, welche eine grobe Klangtextur darstellt. Eine Auswertung dieser in Bezug auf die Zeit findet in der dritten Ebene statt, bei der versucht wird, Muster (z.B. Grundschlag; siehe Abschnitt 3.1.2.2 Paragraph „Rhythmus“) und redundante Informationen zu extrahieren, welche Vorhersagen ermöglichen können, d.h. Anwendung diverser Algorithmen des *Maschinellen Lernens* (z.B. *HMMs*; siehe Abschnitt 5.2.3). Zum Schluss werden die Ergebnisse noch für das Langzeitlernen klassifiziert, d.h. Gruppierungs-, Regressions- und Klassifikationstechniken, wie z.B. *Künstliche Neuronale Netze (KNN)* oder *Support-Vektor-Maschinen (SVM)* werden angewandt.

5.1 Audiosignalextraktion und -anreicherung

Im Laufe der letzten Jahre haben sich eine Vielzahl von Forschungsprojekten im universitären aber auch kommerziellen Umfeld mit dem Thema der Audiosignalextraktion und der Anreicherung der gewonnenen Parametern (siehe Abschnitt 3.1.1) hin zu höherwertigen Musikmerkmalen (siehe Abschnitt 3.1.2) beschäftigt. Zahlreiche *Audiosignalanalyserahmenwerke* wurden mit der Absicht entwickelt, diesen Prozess besser zu begleiten und zu beschreiben.

Der Weg der Herleitung von angereicherten Musikmerkmalen ist zwar auf den ersten Blick nicht für den Endanwender von großer Bedeutung, aber besonders für den Vergleich der Ergebnisse verschiedener Systeme. Weiterhin spielt auch die Beschaffenheit des vorhandenen Audiosignals eine wesentliche Rolle, d.h. ob und welcher Form es komprimiert ist.

Eine recht umfangreiche Grundlage wurde mit dem in Abschnitt 4.2.2.1 beschriebenen *MPEG-7-Standard* geschaffen, welcher eine große Zahl an einfachen Audiosignaleigenschaften aber auch zusätzlich angereicherte Musikmerkmale beschreibt. Frei verfügbare Rahmenwerke, wie z.B. *MARSYAS*²[Tza02], *CLAM*³[AAG06] oder *Vamp*⁴, aber auch die Echtzeit-Audio-Programmiersprache *ChucK*[WFC07] haben viele Algorithmen zum Erkennen dieser Basiseigenschaften umgesetzt bzw. ermöglichen es diese umzusetzen und implementieren zusätzlich auch noch weitere Funktionalitäten zum Ermitteln anderer Musikbeschreibungen. Auch kommerziell ausgerichtete Produkte, wie z.B. *mufin audio-id*⁵[muf09] orientieren sich an dem *MPEG-7-Standard*.

Der Extraktions- und Analyseprozess kann mit Hilfe von *Web Services* auf verteilte Systeme ausgelagert und dadurch u.a. vorhandene Analyseergebnisse bei Übereinstimmung von eindeutigen Identifikatoren effektiv wieder benutzt werden. So wurden z.B. Musikanalyseanwendungen der MIR-Gruppe der TU Wien⁶ innerhalb des *MUSLE Network of Excellence*⁷ als *Web Services* öffentlich zugänglich gemacht (siehe [Mus09]). Ein weiterer *Web Service*[The08] beinhaltet u.a. die in Jehan's Doktorarbeit entwickelten Musik-

¹Auch als *Audio-Fingerabdruck* bezeichnet; engl. *Audio Fingerprint*

²*Music Analysis, Retrieval, and SYNthesis for Audio Signals*

³*C++ Library for Audio and Music*

⁴Welches u.a. durch seine Plug-in-Technik auch *MARSYAS* einbinden kann, siehe <http://www.vamp-plugins.org/>

⁵Welches aus der am Fraunhofer Institut für Digitale Medientechnologie Illmenau entwickelten Technologie *AudioID* hervorgegangen ist; siehe [Fra09]

⁶<http://www.ifs.tuwien.ac.at/mir/>

⁷<http://www.muscle-noe.org/>

analysetechniken (siehe [Jeh05]) und wird durch das Unternehmen *EchoNest*⁸ für nicht-kommerzielle Zwecke kostenlos zur Verfügung gestellt.

5.1.1 Audiosignalextraktion - Grundlagen

Bei der Audiosignalanalyse findet i.A. eine diskrete *Abtastung* (engl. *sampling*), auch *Quantisierung* genannt, des kontinuierlichen Audiosignals statt, d.h. es wird dabei in viele kleine *Abschnitte* (engl. *frames* oder auch *samples*) einer bestimmten Dauer (*Abtastrate*) zerlegt, in denen die in diesem Zeitraum vorherrschenden Werte von verschiedenen Eigenschaften zur Beschreibung des Signals in einem sogenannten *Behälter* (engl. *bin*) oder Vektor (mit einer Dimension die der Anzahl der Eigenschaften entspricht) festgehalten werden. Bei diesem Prozess ist die Wahl der Abtastrate bzw. *Abtastfrequenz* entscheidend. Dem *Nyquist-Shannon-Abtasttheorem* folgend sollte die Abtastfrequenz f_s mindestens doppelt so hoch sein wie die höchste abzubildende Spektralkomponente f_a (siehe [Mei05]). Die auch bei der *MP3*-Kodierung häufig verwendete Abtastfrequenz von 44,1 kHz kann so z.B. einen Frequenzbereich von bis zu 20 kHz und somit den gesamten Hörbereich abdecken (siehe Abschnitt 3.1.1.1 Paragraph „Frequenz“).

Neben den *Abschnitten*, deren Größen durch die Abtastrate bestimmt ist, werden unterschiedlich große *Ausschnitte* (auch *Fenster*, engl. *window*) des Signals zur Analyse verwendet. Diese können im Falle eines diskreten Signals mehrere *Abschnitte* enthalten. Tzanetakis unterteilt die *Ausschnitte* in seiner Doktorarbeit [Tza02] in *Analysefenster* und *Texturfenster*.

Analysefenster sind dabei kurze, möglicherweise auch überlappende Zeitausschnitte, welche i.A. so klein gehalten werden sollen, dass die Frequenzcharakteristiken des Magnitudenspektrums stabil sein sollen, z.B. 23 ms (siehe [Tza02]). *Texturfenster* hingegen beschreiben eher größere *Ausschnitte*, z.B. 1 s, welche mehrere *Analysefenster* enthalten und damit auch deren verschiedene Charakteristiken. Daraus können sich dann auch bestimmte *Dauermuster* (siehe Abschnitt 3.1.2.2 Paragraph „Rythmus“) ergeben bzw. Mittelwerte und deren Abweichungen berechnet werden. Die *Klangtextur* beschreibt demnach mittels statistischer Charakteristiken, d.h. numerischen Merkmalen, die Spektralverteilung komplexer Audiosignale (siehe [Tza02]).

Diese Unterteilung lässt sich auch gut auf das schon am Anfang dieses Kapitels vorgestellte *Musiksignalanalyserahmenwerk* anwenden (siehe Abb. 5.1). Der Bereich „Gehör“ kann mit 10 ms dabei noch zu den *Analysefenstern* gezählt werden und die Bereiche „Merkmalsextraktion“, „Kurzzeitgedächtnis“ und „Langzeitgedächtnis“ mit 100 ms, 1 s bzw. mehreren Sekunden zu den *Texturfenstern*.

Des Weiteren können die Fenstertypen auf die von Peeters [Pee04] aufgestellte zeitbezogene Unterteilung der Deskriptoren angewandt werden. *Globale Deskriptoren* haben dabei eine Gültigkeit für das gesamte Signal (z.B. Mittelwerte oder Abweichungen), welches entweder die Länge eines Texturfensters darstellt oder eine Unterteilung in nicht überlappende Ereignisse, z.B. Einschwingphase, Ausschwingphase. *Instante Deskriptoren* werden hingegen pro Analysefenster berechnet (hier mit einer durchschnittlichen Länge von 60 ms angegeben), d.h. zu einem gegebenen Zeitpunkt, z.B. der Spektralschwerpunkt, welcher sich im Zeitverlauf ändern kann (siehe Abschnitt 5.1.4.2).

⁸<http://the.echonest.com/>

5.1.2 Zeit-Frequenz-Transformation

Die Umwandlung eines zeitdiskreten Signals in den Frequenzbereich heißt i.A. *diskrete Zeit-Frequenztransformation* und gibt das Spektrum (siehe Abschnitt 3.1.1.1) des abgetasteten Signals an. Bekannte und gebräuchliche Techniken sind dabei die Anwendung einer *Diskreten Fourier-Transformation (DFT)* oder einer *Diskreten Wavelet-Transformation (DWT)* mit der Zeit als Startfunktion und der Frequenz als Zielfunktion. Die Rücktransformation ist i.A. als *Synthese (Fourier-Synthese, Wavelet-Synthese)* bekannt.

Die am häufigsten verwendete Technik zur Musiksignalanalyse ist dabei wohl die *Diskrete Fourier-Transformation* mit dem effizienten Algorithmus der *Schnellen Fourier-Transformation* (engl. *Fast Fourier-Transformation, FFT*) und dem zugehörigen Umkehralgorithmus *Inverse FFT (iFFT)*. Eine lineare und konstante zeitliche Begrenzung dieser Transformation ist mit Hilfe der *Kurzzeit-Fourier-Transformation* (engl. *Short-Time Fourier-Transformation; STFT*) möglich, welche das Signal in die bereits beschriebenen Fenster unterteilt. Diese spezielle *DFT* ist dabei die verbreitetste Form in der Audiosignal-extraktion⁹. Die Parameter *Fenstergröße*, *Sprunggröße* (engl. *hop size*) und eine bestimmte *Fensterfunktion* (z.B. *von-Hann-Funktion* oder *Hamming-Funktion*) wirken dabei auf die *STFT* ein. Die *Sprunggröße* (z.B. 256 Abschnitte) ist dabei häufig halb so groß wie die *Fenstergröße* (z.B. 512 Abschnitte), womit sich die *Fenster* überlagern. Dies hängt i.A. aber vom Verwendungszweck der Analyse ab (siehe [IS05]). Eine *Fensterfunktion* wird zur Verminderung des *Leck-Effektes*, d.h. einer Glättung des Frequenzspektrums, eingesetzt, um durch eine Abrundung des Signals, d.h. eine Schwächung der Amplitude an Anfang und Ende des *Ausschnittes*, eine künstliche Periodizität zu schaffen.

Als Alternative zur *STFT* gilt die *Wavelet-Transformation*, welche den *STFT-Nachteile* entgegenwirkt. Die konstante Zeitauflösung von *STFT* wird dabei durch eine hohe Zeitauf-lösung und eine niedrige Frequenzauflösung für hohe Frequenzen, und eine niedrige Zeitauf-lösung mit einer hohen Frequenzauflösung für niedrige Frequenzen verbessert (siehe [Tza02]). Diese Anpassung ähnelt auch sehr dem Verhalten des menschlichen Hörsystems, dessen unterschiedliche Frequenzwahrnehmung grob durch die variable Fenster-größen modelliert werden können (siehe Abschnitt 3.1.1.2 Paragraph „Tonheit“).

Eine direkt an den Anforderungen der Musiksignalanalyse aufbauende Zeit-Frequenz-Transformation ist die *Variable Resolution Transform*¹⁰ (*VRT*; siehe [Par07]). Diese Technik wurde durch die *Wavelet-* und *Constant-Q-Transformation*[Bro91] inspiriert. *VRT* ist eine Art Filterbank, welche sich entlang einer logarithmischen Frequenzskala auf-reiht und versucht, eine relativ konstante Frequenzortung in allen Frequenzbereichen zu ermöglichen. Dafür wird eine sich verändernde *Mutter-Wavelet-Funktion* eingesetzt, welche durch zwei Parameter linear und exponentiell in der *Fenstergröße* angepasst werden kann (anstatt einfach nur skaliert und verschoben, wie bei der *Wavelet-Transformation*¹¹; siehe [Par07]). Damit kann der dramatische Frequenzauflösungsabfall von *Wavelet-Trans-formationen* in hohen Frequenzbereichen kompensiert werden und es können z.B. auch höhere Harmonische von hohen Tonhöhen unterschieden werden (siehe Abschnitt 5.1.5.3). Eine Erfassung aller Oktaven mit einer gleichbleibenden Frequenzauflösung ist durch *VRT* möglich (vgl. [PC09]).

⁹*STFT* gilt als *de-facto*-Standard in der Musiksignalanalyse [Par07]

¹⁰Auf deutsch heißt dies so viel wie „Veränderliche-Auflösung-Transformation“

¹¹Deshalb ist *VRT* auch „nur“ eine *Wavelet*-ähnliche Transformation

5.1.3 Perzeptionsmodelle

Nach der Zerlegung des Signals in seine Frequenzbestandteile können diese auf Modelle des Hörsystems angewandt werden. Dabei ergeben sich grob die drei folgenden Schritte

- Außen- und Mittelohrfilterung
- Zerlegung in Frequenzgruppen
- Auswertung der Maskierung

Außen- und Mittelohr-Filterung Bei der *Außen- und Mittelohr-Filterung* wird eine Funktion auf die Frequenzen angewandt, welche deren Amplituden verändert, d.h. z.B. eine Anhebung der Frequenzen zwischen ca. 2 und 5 kHz (siehe [Jeh05]). Dabei werden die Intensitäten oft zuerst in die Dezibel-Skala umgewandelt (siehe Abschnitt 3.1.1.1 Paragraph „Intensität“). Diese Anpassung (Filterung) soll der Signalverarbeitung des menschlichen Ohres in diesem Bereich nachempfunden sein (siehe Abschnitt 3.1.1.2).

Frequenzgruppenzerlegung Bei der Zerlegung in Frequenzgruppen, welche die Frequenzanalyse im Bereich des Innenohres wiedergeben soll (siehe Einleitung Abschnitt 3.1.1.2), können verschiedene Modelle angewandt werden. Die Intensität, d.h. indirekt das Lautheitsempfinden (siehe Abschnitt 3.1.1.2 Paragraph „Lautheit“), der Frequenzgruppen wird dabei jeweils für die Frequenzgruppe aufsummiert. Das erste bedeutende Modell ist die *Mel-Skala* (siehe Abschnitt 3.1.1.2 Paragraph „Tonheit“), welche besonders bei der Spracherkennung genutzt wird (vgl. [Pee04]). Es ist grob eine logarithmische Skala der linearen Frequenzskala und lässt sich folgendermaßen berechnen (wobei f die Frequenz in Hz ist; siehe [O’S87]):

$$Mel(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (5.1)$$

Die Anzahl der Unterteilungen in Frequenzgruppen und deren Größe ist dabei nicht fest vorgeschrieben. Es gibt z.B. gleich große Aufteilungen in 20[Pee04] oder 40[Log00] Frequenzbänder.

Die *Bark-Skala* (siehe Einleitung Abschnitt 3.1.1.2) ist ein anderes Modell der Frequenzwahrnehmung, welches sich besser an das Verhalten des menschliche Hörsystem annähern soll [Pee04]. Die Umwandlung der Frequenzskala in die *Bark-Skala*, lässt sich durch folgende Formel beschreiben (siehe [FZ07]):

$$z/Bark = 12 \cdot \arctan(0.76f/kHz) + 3.5 \cdot \arctan(f/7.5kHz)^2 \quad (5.2)$$

Letztendlich lassen sich die errechneten Werte auch leicht zwischen den Skalen umrechnen, da diesen in einem linearen Verhältnis zueinander stehen (siehe Abschnitt 3.1.1.2 Paragraph „Tonheit“).

Maskierungsauswertung Nach dieser Reduktion auf Frequenzgruppen lassen sich Funktionen anwenden, welche die *Frequenz-* und *Zeitmaskierung* bei der Wahrnehmung simulieren. Die simultane Maskierung von Klängen findet auch über die Grenzen der Frequenzbänder hinweg statt und lässt sich einfach durch eine *Dreiecksfunktion* mit $+25dB/Bark$ und $-10dB/Bark$ annähern (siehe [Jeh05]). Tiefere Frequenzen haben dabei einen

größeren Einfluss auf höhere Frequenzen. Komplexere *Verteilungsfunktionen* finden auch in der verlustreichen Signalkodierung Anwendung und sind dort ein gutes Qualitätsmerkmal.

Die *Zeitmaskierung* steht dabei mit dem Merkmal Tondauer (siehe Abschnitt 3.1.2.1 Paragraph „Tondauer“) in Verbindung, d.h. auch mit der Hörsinnesempfindung Schwankungsstärke (siehe Abschnitt 3.1.1.2 Paragraph „Schwankungsstärke und Grobheit“), und bezieht sich dabei im Besonderen auf die *Nachmaskierung*, da diese einen relativ großen Zeitraum von ca. 200 ms einnimmt¹². Durch Anwendung von *Fensterfunktionen* (siehe Abschnitt 5.1.2) lässt sich dieser Glättungseffekt gut abbilden (vgl. [Jeh05]).

5.1.4 Statistiken

Werte der Zeit- und Frequenzanalyse von diskreten Audiosignalen können mit Hilfe statistischer Berechnungen die Dimension der Merkmalsvektoren drastisch reduzieren und darüber hinaus auch noch brauchbare Informationen zur Merkmalsanreicherung liefern. Diese Statistiken können entweder direkt auf dem Audiosignal (siehe Abschnitt 5.1.4.1), Frequenzspektrum (siehe Abschnitt 5.1.4.2) oder darauf angewandten Perzeptionsmodellen berechnet werden (siehe Abschnitt 5.1.4.3).

Im Allgemeinen können von den errechneten Werten auch noch weitere Mittelwerte und Abweichungen über *Texturfenster* oder noch größere Zeitbereiche (z.B. ein ganzes Musikstück) berechnet werden.

5.1.4.1 Direkte Statistiken

Direkte Statistiken können direkt aus dem Signal gewonnen werden und benötigen deshalb keine weiteren Vorberechnungen. Im Folgenden werden die wichtigsten davon kurz erklärt.

Nulldurchgangsrage Besonders im Bereich der Spracherkennung und des *Musikinformationsretrievals* ist der Parameter *Nulldurchgangsrage* (engl. *Zero Crossing Rate*, *ZCR*) von Bedeutung. Er gibt die Anzahl der Nulllinienüberschreitungen des Signales an, z.B. die Anzahl der Wechsel von positiven zu negativen Werten innerhalb einer Sekunde (d.h. einem Texturfenster¹³). Mit der *Nulldurchgangsrage* können dominante Frequenzen (niedriger Wert) aber auch das Rauschen (hoher Wert) eines Signales ermittelt werden (siehe [Lid06] und [Pee04]). Somit dient dieser Wert als Basisabgrenzer zwischen Sprache und Musik und trägt auch zur Genreeinordnung von Musik bei.

Zum Beispiel hat *Heavy-Metal*-Musik eine höhere *Nulldurchgangsrage* als klassische Musik (verursacht durch die verzerrten Gitarren und einem schnellen Rhythmus des Schlagzeuges [Tza02]). Eine reine *Nulldurchgangsrage* (d.h. ohne Rauschen) steht eng im Zusammenhang mit dem *Spektralschwerpunkt* (siehe Abschnitt 5.1.4.2 Paragraph „Spektralschwerpunkt“).

Energieeffektivwert Von der Intensität abgeleitet, lässt sich beim Abtasten von Audiosignalen deren *Energieeffektivwert* (engl. *Root Mean Square Energy*; deshalb auch *RMS-Energie*) ermitteln, welcher die Quadratwurzel des Mittelwertes des Quadrates aller

¹²Im Gegensatz zur *Vormaskierung* mit ca. 20 ms

¹³kann aber auch pro *Analysefenster* berechnet und danach gemittelt werden

Amplituden in einem *Ausschnitt* ist (*quadratisches Mittel*). Dieser Wert ist eine gute Charakterisierung für die Lautstärke (siehe Abschnitt 3.1.1.2 Paragraph „Lautheit“) in diesem Zeitfenster [Lid06] und auch eine gute Grundlage für die Audioereigniserkennung, Segmentierung oder Tempo- bzw. Grundschlagschätzung (siehe Abschnitt 3.1.2.2 Paragraph „Rhythmus“). Im Speziellen können darauf weitere Beschreibung berechnet werden, wie z.B. *Einschwingzeit* (siehe Abschnitt 3.1.1.2 Paragraph „Einschwingzeit“), oder die Länge und Neigungsgrade der Ankling- und Abklingphase, welche für die Klangfarbenbestimmung (siehe Abschnitt 3.1.2.1 Paragraph „Klangfarbe“) genutzt werden können (vgl. [Pee04]).

Niedrigenergie rate Eng verwandt mit dem *Energieeffektivwert* ist die *Niedrigenergie rate* (engl. *Low Energy Rate*). Sie trägt auch zur Einschätzung der zuletzt genannten Musikmerkmale bei. Dieses Maß gibt die Prozentzahl der Analysefenster an, welche weniger *RMS-Energie* beinhalten als der durchschnittliche Energiewert (siehe [Lid06]). Somit wird die *Niedrigenergie rate* über einem *Texturfenster* ermittelt. Ein höherer Wert lässt hierbei auf eine höhere Dynamik schließen (siehe Abschnitt 3.1.2.2 Paragraph „Musikvortragsbezeichnungen“).

Musiksignale haben i.A. eine geringere *Niedrigenergie rate* als Sprachsignale [Tza02].

5.1.4.2 Spektralstatistiken

Diese Statistiken werden nach der Zeit-Frequenz-Transformation auf dem Spektrum des jeweiligen Fensters angewandt bzw. direkt auf den vorhandenen Merkmalsvektoren bei vor-kodierten Audiosignalen, wie z.B. *MP3*¹⁴ (siehe [Tza02]). Im Folgenden werden nun die davon wichtigsten kurz vorgestellt.

Spektralfluss Der *Spektralfluss* (engl. *Spectral Flux*) gibt die lokale Veränderungsrate (Fluktuation) des Spektrums durch Vergleichen von kleinen aufeinanderfolgenden *Analysefenstern* an. Dieses Maß wird u.a. zur Ermittlung von Klangfarbe und *Einschwingzeit* (siehe Abschnitt 3.1.2.1 Paragraph „Klangfarbe“ und Abschnitt 3.1.1.2 Paragraph „Einschwingzeit“) genutzt (siehe [Lid06]).

Es lässt sich z.B. als Quadrat der Differenz zwischen normalisierten Magnituden von aufeinanderfolgenden Spektralverteilungen berechnen (siehe [Tza02]).

Spektralschwerpunkt Der *Spektralschwerpunkt* (engl. *Spectral Centroid*) ist das Zentrum der Anziehungskraft, also in diesem Fall das Zentrum des Frequenzspektrums. Er gibt die Frequenz an, bei der die Energie aller Frequenzen unter dieser Frequenz gleich der Energie aller Frequenzen über dieser Frequenz ist. Diese Größe dient zur Berechnung der *Helligkeit* (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“) und der generellen Spektralform (siehe [Lid06]), aber auch zur Bestimmung der Klangfarbe von Instrumenten (siehe [Tza02]). Je höher der *Spektralschwerpunkt*, umso heller klingt der Ton. Dieses Maß ist ähnlich der Hörsinnesempfindung *Schärfe* (siehe Abschnitt 5.1.4.3 Paragraph „Schärfe“ und [FZ07]).

Es lässt sich als gewichtetes arithmetisches Mittel berechnen, wobei die Magnituden der vorhandenen Frequenzen des Ausschnittes die Gewichte sind (siehe [Tza02]).

¹⁴Dies spart Berechnungszeit, da das Signal schon in 32 Frequenzbänder unterteilt ist

Spektral-Rolloff Der *Spektral-Rolloff* (engl. *Spectral Rolloff*) ist ein weiteres Maß zur Spektralform. Er gibt die α -ige Überschreitungswahrscheinlichkeit der Spektralverteilung an, d.h. bis zu welcher Frequenz sich die meiste Signalenergie befindet. Der Wert α liegt dabei üblicherweise zwischen 85 und 95 (vgl. [Tza02], [Jeh05] und [Pee04]). Diese Größe ist deshalb ein Maß der Schiefe der Spektralform und zeigt an, wie viel Signalenergie sich auf tiefe Frequenzen konzentriert [Jeh05]. Der *Spektral-Rolloff* steht auch in Verbindung mit der Harmonie/Rausch-Grenzfrequenz, der sogenannten *Nyquist-Frequenz* (Abtastfrequenz; siehe Kapitelanfang).

5.1.4.3 Perzeptionsstatistiken

Unter Perzeptionsstatistiken werden Beschreibungen zusammengefasst, welche auf Basis von Perzeptionsmodellen (siehe Abschnitt 5.1.3) berechnet werden.

Mel-Frequenz-Cepstrum-Koeffizienten Die *Mel-Frequenz-Cepstrum-Koeffizienten* (engl. *Mel-Frequency-Cepstrum-Coefficients*; *MFCC*) repräsentieren die Spektraloberfläche eines Audiosignals mit wenigen Koeffizienten [Pee04]. Im Allgemeinen wird ein *Cepstrum* auf einen logarithmischen Frequenzspektrum mit Hilfe einer weiteren *FFT* der Dezibel-Skala berechnet (siehe Abschnitt 5.1.2). Im Fall der *MFCC*-Berechnung wird die Abbildung des Frequenzspektrums auf die *Mel-Skala* vorgenommen (siehe Abschnitt 5.1.3). Als Transformationstechnik wird häufig eine *Diskrete Kosinus-Transformation*¹⁵ (engl. *Discret Cosine Transformation*; *DCT*) angewandt (anstatt eine rechenintensivere *FFT*), da die Werte des Eingangssignals oft reell und symmetrisch sind [Whi05]. Der Betrag der *Frequenzbänder* der *Mel-Skala* gibt das Maximum der berechenbaren *MFCCs* an. Bei der Spracherkennung sind dies typischerweise 13, für die Musikmerkmalsanalyse reichen aber schon die ersten 5 (z.B. als Grundlage zur automatischen Musikgenreklassifikation [Tza02]) aus. Dennoch variiert die Anzahl der Koeffizienten stark und ist vom jeweiligen Verwendungszweck abhängig, z.B. 19 für Musikähnlichkeit (siehe [LS01]). Diese kompakte Form der Spektraloberflächenbeschreibung kann zur Klangfarbenbeschreibung genutzt (siehe [Tza02]) aber auch in *Gaussian-Mixture-Modellen* (*GMM*; siehe Abschnitt 5.2.3) aggregiert werden, welche als Grundlage für Kategorisierungen von z.B. Musikstücken oder Künstlern genutzt werden kann (siehe Abschnitt 3.2), d.h. Ähnlichkeitsberechnungen (z.B. Berechnung der *Kullback-Leibler-Divergenz*[Rai08a]) darauf vorgenommen werden können (siehe [DSPK08]). Generell wird die *MFCC*-Charakteristik häufig für diese Berechnungen als Datenquelle genutzt.

Schärfe Die *Schärfe* ist das perzeptuelle Äquivalent zum *Spektralschwerpunkt* (siehe Abschnitt 5.1.4.2 Paragraph „Spektralschwerpunkt“) und lässt sich aus den einzelnen Lautheitswerten der einzelnen Frequenzgruppen der *Bark-Skala* berechnen. Dieses Maß gehört in den Bereich der Klangfarbenrepräsentation und wird von der Mittelfrequenz und dem Spektralumfang des *Frequenzbandes* beeinflusst (siehe [FZ07]). Zwickler et al. haben für dieses Maß sogar eine eigene Größe entworfen, welche *acum* heißt. Wobei 1 acum als ein frequenzgruppenbreites Schmalbandrauschen bei einer Mittelfrequenz von 1 kHz und einer Amplitude von 60 dB definiert ist.

¹⁵Wobei die Dekorrelation mit einer *DCT* grob gleiche Ergebnisse liefert wie eine *Hauptkomponentenanalyse* (engl. *Principal Component Analysis*; *PCA*) [Log00]

5.1.5 Histogramme

Histogramme (z.B. Spektrogramme, siehe Abschnitt 3.1.1.1 Paragraph „Spektrum“) sind i.A. eine Repräsentationsform von (angereicherten) statistischen Informationen, deren Werte in Form von Rechtecken dargestellt werden und deren Unterteilung durch gleichgroße numerische Werte erfolgt. Sie können gut zur Repräsentation von Rhythmus-, Tonhöhen- und Klangfarbenmerkmalen genutzt werden, wovon sich auch die Überschriften der folgenden Unterkapitel ableiten. Neben der Bestimmung von angereicherten Audiosignaleigenschaften (siehe Abschnitt 3.1.2) können diese Histogramme auch direkt für Ähnlichkeitsberechnungen genutzt werden. Dabei werden sie entweder in unterschiedlichen Varianten zusammen linear kombiniert, oder zum Training für Algorithmen des *Maschinellen Lernens* (z.B. *KNNs*) genutzt (vgl. [PC09]).

5.1.5.1 Histogramme zur Rhythmusrepräsentation

Rhythmusrepräsentationen können in erster Linie für die Grundsclagerkennung genutzt werden, aber auch zur Erkennung komplexer Rhythmen. Damit sind auch Eigenschaften gemeint, wie z.B. die Beziehung des Grundsclages zu den nicht voll betonten *Tatums* bzw. Vielfache des *Tatums*, d.h. die relative Stärke dieser zum Grundsclag (siehe Metrum; Abschnitt 3.1.2.2 Paragraph „Rhythmus“).

Im Allgemeinen ähneln sich die Berechnungsschritte der Algorithmen zur Berechnung von Histogrammen der Rhythmus- und Tonhöhenrepräsentation. Der Unterschied liegt grob nur in der Wahl des Ausschnittes. Zur Erkennung von Grundsclagen sollten i.A. eher größere *Ausschnitte*¹⁶ (z.B. 0,5 – 3 s) verwendet werden (um auch langsame Rhythmen oder Grundscläge zu erkennen) und zur Tonhöhenerkennung eher kleine (z.B. 2 – 50 ms).

Eine grobe Berechnungsfolge zur Rhythmus-(/Grundsclag-)Erkennung setzt sich aus den folgenden drei Schritten zusammen (siehe [Tza02]):

1. Zerlegung in Frequenzgruppen, wobei die Anzahl der verwendeten Frequenzgruppen unterschiedlich ist, z.B. 12 (*Oktaven-Frequenzband-Zerlegung*), 24 (*Bark-Frequenzbänder*), 36 (gebräuchliche *Mel-Frequenzband-Zerlegung*)
2. Hüllkurven-Extraktion, d.h. die Intensitäten der Frequenzgruppen im Laufe der Zeit
3. Anwendung eines *Periodizitätserkennungsalgorithmus*, d.h. zu welchen Zeitpunkten die Signalhüllkurve am ähnlichsten zu sich selbst ist

Tzanetakis entwickelte in seiner Doktorarbeit [Tza02] ein sogenanntes *Grundsclag-Histogramm*¹⁷ (engl. *Beat Histogram*), welchem andere Ansätze, wie z.B. Lidy [Lid06], mit seinem *Rhythmus-Histogramm*, Whiteman [Whi05], mit seinem *Beatogram*, oder Paradziniets [Par07] mit seinem *2D-Grundsclag-Histogramm*¹⁸, folgten. Ähnlich ist auch das *tatumbasierte*¹⁹ *Tatum-Raster* (engl. *Tatum Grid* [Sep01]), welches u.a. von Jehan verbessert wurde (siehe [Jeh05]).

Zur Periodizitätserkennung werden oft sogenannte *Autokorrelationsfunktionen* eingesetzt,

¹⁶Zur *Tatum*-Erkennung werden dementsprechend kleinere Ausschnitte genutzt

¹⁷Welches im Gegensatz zu den anderen vorgestellten Ansätzen eine *Wavelet-Transformation* nutzt

¹⁸Welches auf *VRT* basiert und mittels Schwellwerten versucht Anomalien, z.B. Lautstärkeschwankungen, auszugleichen; siehe [Par07]

¹⁹Wie auch das *2D-Grundsclag-Histogramm* zur *Tatum*-Erkennung genutzt werden kann

welche genau die Ähnlichkeit eines Signals zu sich selbst beschreiben. Die davon resultierenden Maxima geben dabei die unterschiedlichen Perioden wieder. Eine unterschiedlich große Anzahl der höchsten (dominanten) Maxima wird davon mit in die Darstellung des Histogrammes zur Rhythmusrepräsentation übernommen. Hohe Amplituden bei diesen Maxima zeigen starke Grundschläge bzw. *Tatums* auf.

Einen etwas veränderten Ansatz hat Pampalk[Pam01, Pam06] mit seinen *Rhythmus-Mustern*²⁰ (engl. *Rhythm Pattern*) gewählt, welcher von Pflugfelder[Pfl08] umfangreich auf den Einsatz zur automatischen Musikgenreklassifikation evaluiert wurde und woraus sich Lidy's *Rhythmus-Histogramme* mit ableiten (siehe [Lid06]). Ein *Rhythmus-Muster* ist grob eine Matrix von Schwankungen auf *Frequenzbändern* [Lid06]. Hierbei wird die Modulationsfrequenz auf den durch die *Bark-Skala* bestimmten *Frequenzbändern* auf Basis der Lautheitswerte (in Sone) mit Hilfe einer *FFT* berechnet (d.h. das *Cepstrum*²¹) und anschließend damit durch die jeweilige Schwankungstärke gewichtet (anstatt der *Autokorrelation*). Bei der Repräsentation werden nur die ersten 60 Modulationsfrequenzen benutzt, welche einen Bereich von 0,17 bis 10 Hz (600 BPM) abdecken und damit gut die in der Musik verwendeten Periodizitäten einschließen [Lid06].

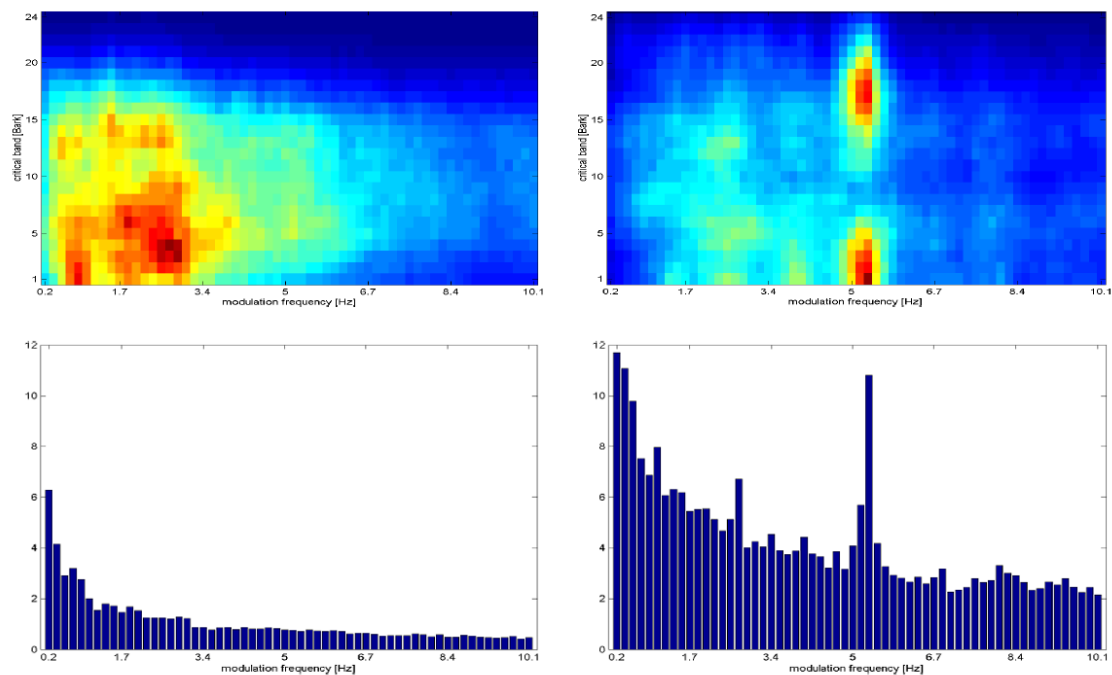


Abbildung 5.2: *Rhythmus-Muster* (oben) und *Rhythmus-Histogramm* (unten) auf Basis von 60 Modulationsfrequenzen [Lid06]

Weitere Statistiken über den Werten der Histogramme zur Rhythmusrepräsentation (Zeitpunkte und Amplituden der Maxima) können Aussagen über Metrum, Takt und *Musikvortragsbezeichnungen* liefern (siehe Abschnitt 3.1.2.2 Paragraph „Rhythmus“ und „Musikvortragsbezeichnungen“). In Abbildung 5.2 sind zwei beispielhafte *Rhythmus-Muster* und *Rhythmus-Histogramme* zu sehen, wobei jeweils das links dargestellte, mit einem Ausschnitt klassischer Musik (mit vielen Instrumenten eines Orchesters), weniger do-

²⁰In dieser Arbeit hießen sie noch *Schwankungsmuster*, engl. *Fluctuation Pattern*

²¹Deshalb können hier auch *MFCCs* genutzt werden; siehe [Pam06]

minante Maxima (Grunds Schlagstärke²²) aufweist als das rechts dargestellte, welche von einem Rock-Musikstück erstellt wurden (siehe [Lid06]).

5.1.5.2 Histogramme zur Tonhöhenrepräsentation

Histogramme zur Tonhöhenrepräsentation berechnen sich grob ähnlich wie die im letzten Unterkapitel vorgestellten Histogramme zur Rhythmusberechnung. Die wesentlichen Unterschiede sind nur die verwendeten Größen der *Analysefenster*, die Anzahl der Frequenzgruppen und die Anwendung eines *Mehrfach-Tonhöhenanalysealgorithmus*, welcher auf *polyphone* Musiksignale angewendet werden kann (z.B. [PHC06]). Bekannte Vertreter sind das *Tonhöhen-Histogramm* von Tzanetakis[Tza02] und das *Chromagramm*[BW01], welches u.a. von Jehan[Jeh05] verbessert wurde.

Das *Chromagramm* ähnelt dabei sehr dem *gefalteten Tonhöhen-Histogramm* (engl. *Folded Pitch Histogram*, *FPH*[Tza02]), weil beide die erkannten Tonhöhen in die 12 Tonklassen einordnen (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“). Der Vorteil von Jehan's *Chromagramm* liegt in den durch das *Tatum-Raster* bereits analysierten Segmentierungen, welche als nicht-überlappende (größere) *Analysefenster* zur Tonhöhenerkennung genutzt werden können.

Auseinandergefaltete Tonhöhen-Histogramme (engl. *Unfolded Pitch Histograms*, *UPH* [Tza02]) geben die *Helligkeit* an, d.h. damit können Aussagen über die Tonhöhenreichweite gemacht werden. *Gefaltete Tonhöhen-Histogramme* können hingegen gut zur Chroma-, Tonart-, Akkord- und Harmonie-Bestimmung genutzt werden (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“ und Abschnitt 3.1.2.2 Paragraph „Harmonie“, und [Jeh05]).

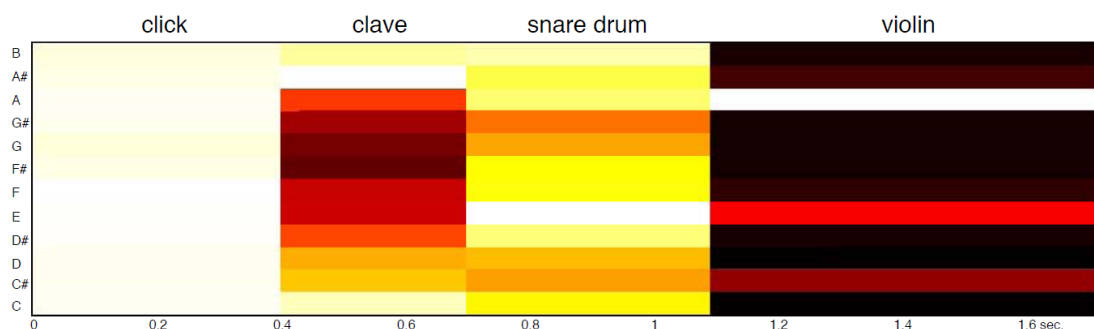


Abbildung 5.3: Eine *Chromagramm* von vier Klängen (Klick, Clave, Snare-Drum, Violine) [Jeh05]

Abbildung 5.3 zeigt ein Beispiel-*Chromagramm* von Jehan, auf dem deutlich die unterschiedlichen Längen der Segmentierungen der einzelnen Klänge zu erkennen sind und auch die ggf. vorherrschenden Tonhöhen, durch eine sehr helle bis weiße Farbgebung bei der dominanten Tonklasse, d.h. einer hohen Intensität in dieser Frequenzgruppe, und dunkleren Einfärbungen bei den anderen Tonklassen. Zum Beispiel wird auf der Violine ein A gespielt oder ein starkes *Rauschen* (siehe Abschnitt 3.1.1.2 Paragraph „Schwankungsstärke und Grobheit“) ist bei dem Klick zu erkennen.

Weiterentwicklungen im Bereich der Histogramme zur Tonhöhenrepräsentation sind die auf *VRT-Spektrogrammen* basierenden Ansätze *Noten-Histogramm* (auch *Notenprofil* genannt) und *Notenfolgen-Histogramm* von Paradzinets (siehe [Par07, PC09]). Auf der einen

²²Woraus auch auf die Betonung für das Metrum Rückschlüsse gezogen werden können

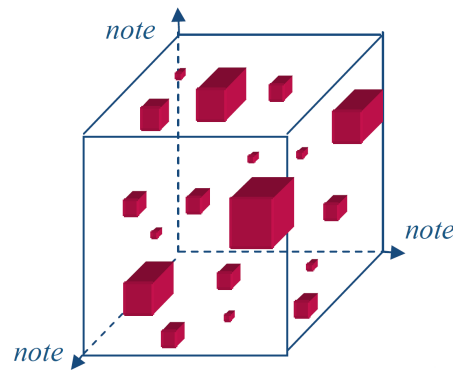


Abbildung 5.4: Ein Beispiel eines *Notenfolgen-Histogrammes* [Par07]

Seite werden bei dem *Notenprofil* die erkannten Noten zur Tonartschätzung genutzt, wodurch ggf. auf bestimmte Stimmungen Rückschlüsse gezogen werden können (siehe Abschnitt 3.1.2.1 Paragraph „Tonhöhe“). Auf der anderen Seite werden bei dem *Notenfolgen-Histogramm* drei notenlange Notenfolgen von dominanten Noten eines *Texturfenster* extrahiert und in einem *3D-Notenfolgen-Histogramm* dargestellt, wobei jede Achse einer Note dieser Notenfolge entspricht und große Quader besonders dominante Notenfolgen repräsentieren (siehe Abbildung 5.4). Dieser Histogrammtyp repräsentiert im Besonderen ein *Audio-Fingerabdruck* des jeweiligen Musikstückes und kann zur Melodie- und Harmonie-Ähnlichkeitsbestimmung genutzt werden [PC09].

5.1.5.3 Histogramme zur Klangfarbenrepräsentation

Dieser Histogrammtyp versucht die Charakteristiken von Klangfarben (siehe Abschnitt 3.1.2.1 Paragraph „Klangfarbe“) zu repräsentieren und hervorzuheben. Hierbei werden besonders die Ergebnisse aus der Psychoakustik als Grundlage genommen, d.h. die frequenzselektiven Kanäle des Hörsystems – die bereits vorgestellten Bark-Frequenzbänder (siehe Abschnitte 3.1.1.2 und 5.1.3).

Deshalb kann hier die Darstellung der *Rhythmus-Muster* (siehe Abschnitt 5.1.5.1) wieder verwendet, aber auch die von Jehan vorgestellten (recht ähnlichen) *Bark-Spektrogramme* [Jeh05] eingesetzt werden. Diese Ansätze vermitteln eine globale Übersicht über vorkommende Klangfarben und können deshalb gut für Musikähnlichkeitsuntersuchungen genutzt werden (siehe [Pfl08]). Weiterhin können auch Muster daraus abgeleitet werden, z.B. hohe Amplituden bei niedrigeren Frequenzbändern lassen ein Bass-Instrument vermuten (siehe [Pam01]).

Detailliertere Informationen über die Klangfarben als Profil von Harmonischen kann hingegen ein *VRT-Spektrogramm* liefern, wovon wiederum ein spezielles *Klangfarben-Histogramm*[Par07] abgeleitet werden kann. Durch die relativ konstante Frequenzauflösung über die Oktaven hinweg können die Harmonischen auch bei höherfrequenten Tonhöhen noch gut unterschieden werden. In Abbildung 5.5) ist ein *VRT-Spektrogramm* eines Pianospiele zu sehen, wobei die verschiedenen Harmonischen der gespielten Töne (graue Absätze in höheren Frequenzbereichen) zu erkennen sind (auch wenn sie nicht so eine hohe Intensität wie die Grundfrequenz (weißen Absätze) aufweisen können). Dies verdeutlicht auch die empfohlene Parametrisierung der *VRT*, wobei pro Note grob 10 Frequenzbehälter für eine feingranulare Separation genutzt werden (siehe [Par07]). Das Klangfarben-Histogramm ergibt sich dann aus den jeweils erkannten Noten und den rela-

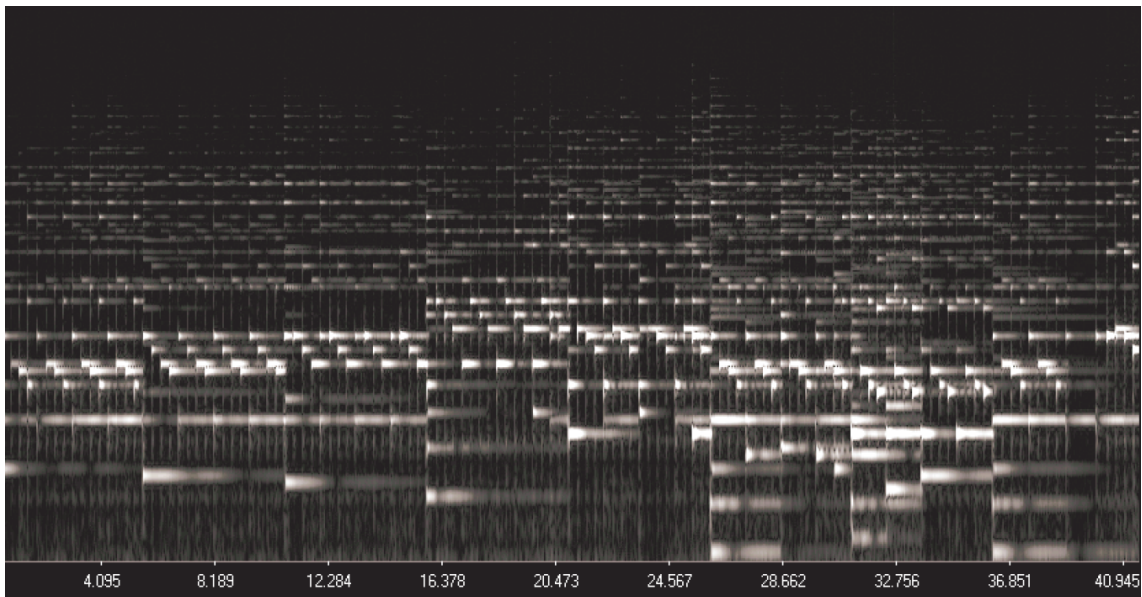


Abbildung 5.5: Ein VRT-Spektrogramm (mit einer logarithmischen Frequenzskala) eines Musikstückauszuges, bei dem nur ein Piano gespielt wird [Par07]

tiven Amplituden der zugehörigen Harmonischen (als reduzierter Zahlenwert pro Note), welche so miteinander verglichen werden können.

5.1.6 Audiosignalanreicherung

Die extrahierten Eigenschaften, Statistiken und zusammengefassten Repräsentationen (siehe vorheriger Abschnitt) werden letztendlich auf unterschiedliche Weise in Form von kompakten Merkmalsvektoren zusammengefügt, welche in ihrer Beschreibungscharakteristik auch als *musikalische Fingerabdrücke* genutzt werden können und von verschiedenen *Musikmetadatenstandards* und *-diensten* eingesetzt werden (siehe Kapitel 4). Zum Beispiel wird bei *MusicBrainz*, *MusicDNS* von der Firma *MusicIP*²³ genutzt, oder eigene Zusammenstellungen wie bei dem *mufin audioid*-Dienst oder *Last.fm* (siehe Abschnitt 4.3). Die Größe der *Audio-Fingerabdrücke* beläuft sich dabei auf ca. 30[Jeh05]-bis 42[Tza02]-dimensionale Merkmalsvektoren, wobei diese oft eine Kombination aus rhythmus-, klangfarben- und tonhöhen-beschreibenden Merkmalen ist (vgl. [Jeh05] und [Tza02]).

Generell werden verschiedene Anforderungen an einen idealen *musikalischen Fingerabdruck* gestellt, welche im Folgenden kurz vorgestellt werden sollen (nach [Can06]):

- Robuste und korrekte Musikstückidentifikation unabhängig vom Grad der Kompression oder des Verlustes, und des Audiodateiformates
- Schnelle Musikstückerkennung innerhalb weniger Sekunden
- Effiziente Berechnung des *musikalischen Fingerabdruckes*
- So kompakt wie möglich

²³<http://www.musicip.com/>

Diese Ansprüche zu erfüllen ist eine große Herausforderung, welcher sich u.a. Cano in seiner Doktorarbeit[Can06] gewidmet hat. Dabei hat er ein komplettes *Audio-Fingerabdruck-Rahmenwerk* entwickelt, welches auch zum *Musikinformatonsretrieval* genutzt werden kann. Ergebnisse dieser und anderer Arbeiten (u.a. [Her08]) sind mit in das Musiktechnologieunternehmen *BMAT*²⁴ eingeflossen, welches u.a. Dienste zur Überwachung und Analyse von Radio-, TV- und Internetmusik auf Basis von *musikalischen Fingerabdrücken* anbietet, z.B. auch die Erkennung von Cover-Versionen von Musikstücken (siehe [BMA09]).

Im Allgemeinen können auf den *akustischen Metadaten* auch diverse Ähnlichkeitsberechnungen angewandt werden, welche verschiedene Strukturierungstechniken (z.B. lineare Kombination oder *Dynamische Programmierung*; siehe Abschnitt 5.2.3.2) als Grundlage nutzen. Die daraus gewonnenen Muster können wiederum als Eingabe für diverse Techniken des *Maschinellen Lernens* genutzt werden. Diese Techniken werden kurz im nächsten Unterkapitel vorgestellt.

5.2 Metadatenanalyse und -anreicherung

Die Metadatenanalyse und -anreicherung als höherwertige Form der Merkmalsextraktion beschäftigt sich grob mit den vier folgenden Bereichen:

1. Auslesen von *formatgebundenen Spezifikationen* (siehe Abschnitt 4.2.1)
2. Auffinden und Auslesen von zusätzlichen *redaktionellen* und *kulturellen Metadaten*
3. Klassifikation und Zuordnung von *angereicherten Audiosignaleigenschaften* (siehe Abschnitt 3.1.2) und *taxonomiebasierten Entitäten* (siehe Abschnitt 3.2) auf Grundlage der drei *Musikmetadattentypen* (siehe Abschnitt 4.1 Paragraph „Metadattentypen“)
4. Hinterlegen der gewonnenen Informationen und Verbindungen in einem geeigneten Musikmetadatenstandard

Nur eine gute und sinnvolle Kombination der verschiedenen *Metadatenfelder* der drei *Musikmetadattentypen* kann zu einer sich dem *Ideal* annähernden Grundlage an Informationen führen, welche für die verschiedenen Anwendungsfälle im privaten Umfeld (siehe Kapitel 2) auf unterschiedliche Weise genutzt werden kann. Das *Ideal* definiert sich hierbei als die Menge an Informationen, die notwendig ist um alle existierenden Anwendungsfälle zur Verwaltung von Musiksammlungen im privaten Umfeld zu bedienen. Die nächsten Abschnitte beschäftigen sich daher mit den Anforderungen, die an die einzelnen, oben beschriebenen Bereiche gestellt werden, um sich diesem Ziel anzunähern.

5.2.1 Metadatenextraktion von formatgebundenen Spezifikationen

Die in Abschnitt 4.2.1 beschriebenen *formatgebundenen Spezifikationen* sollen zur Metadatenanalyse und -anreicherung vollständig ausgelesen werden und in ein einheitliches

²⁴*Barcelona Music & Audio Technologies*; siehe <http://www.bmat.com/>

Format überführt werden. Dieses Format kann z.B. die in Kapitel 4 empfohlene *Music Ontology* sein, welche die *Metadatenfeldergruppierungen*, wie in [CGVD08] demonstriert, vereinen und darüber hinaus auch auf einer unterschiedlichen Granularität repräsentieren kann. Neben dem Auslesen der jeweiligen *Metadatenfelder* innerhalb der Spezifikation, gehört auch eine Erfassung der allgemeinen Eigenschaften, d.h. insbesondere *administrativen Metadaten*, eines Musikdokumentes zu dem in diesem Abschnitt beschriebenen Aufgabenbereich.

Zwei wesentliche Anforderungen sollte eine gute Metadatenextraktionskomponente für *formatgebundene Spezifikationen* erfüllen:

1. Sie soll die Menge an gebräuchlichen Audiosignalformaten (mit ihren Datenbereichen für Metadaten) abdecken und erweiterbar für neu eingeführte *formatgebundene Spezifikationen* sein.
2. Sie soll die Menge an gebräuchlichen *Metadatenfeldern* der jeweiligen *formatgebundenen Spezifikationen* abdecken und erweiterbar für zusätzliche Beschreibungen sein.

Der Vorteil in der Nutzung der *Music Ontology* liegt klar darin, dass dieses *Musikmetadatenformat* genau diese Anforderungen erfüllen kann und darüber hinaus die semantischen Zusammenhänge der einzelnen Eigenschaften so einfach wie möglich modelliert und dabei auf die Wiederverwendung gebräuchlicher Ontologien achtet. Falls ein neu eingeführtes Audiosignalformat oder *Metadatenfeld* nicht durch die derzeit existierende Spezifikation der *Music Ontology* abgedeckt werden kann, kann die Ontologie einfach durch neue Konzepte und Eigenschaften erweitert werden. Generell wurde aber bei der derzeitigen Modellierung der *Music Ontology* aber darauf geachtet, so viel wie möglich der vorhandenen und für die verschiedenen Anwendungsfälle sinnvollen Konzepte einzubringen (vgl. [CGVD08]).

Grundsätzlich werden für jede Programmiersprache zahlreiche Rahmenwerke angeboten, welche ein oder mehrere *formatgebundene Musikmetadatenpezifikationen* verarbeiten können. So z.B. die *Jaudiotagger*-Bibliothek[JTh07], welche *Java*-basiert ist und die u.a. die Formate *ID3*, *Vorbis Comments* verarbeiten kann (siehe [JTh07]). Auf dieser Bibliothek beruht auch der *Mp3FileExtractor* des *Aperture*-Rahmenwerkes²⁵, welches die ausgelesenen Informationen gleich in einen *semantischen Graphen* einer *RDF*-basierten Ontologie, die *NEPOMUK ID3 Ontology*²⁶, umwandelt (siehe [Sau09]).

Auf der anderen Seite gibt es auch Rahmenwerke, welche eine Unterstützung für mehrere Programmiersprachen bereitstellen. Beispielhaft soll hier nur die *TagLib*-Audiometadatenbibliothek[whe08] erwähnt werden, welche für die Programmiersprachen *C*, *Perl*, *Python* und *Ruby* derzeit zur Verfügung steht und die u.a. die Formate *ID3* und *Vorbis Comments* behandeln kann (siehe [whe08]).

Web Services können auch noch diese Aufgabe mit übernehmen, so z.B. die schon in Abschnitt 5.1 zur Audiosignalextraktion vorgestellte *Echo Nest API* (siehe [The09]).

Das Hauptproblem bei der Metadatenextraktion von *formatgebundenen Spezifikationen* besteht darin, dass die *redaktionellen Eigenschaften*²⁷ oft nur unzureichend oder gar nicht

²⁵<http://aperture.sourceforge.net/index.html>

²⁶<http://www.semanticdesktop.org/ontologies/2007/05/10/nid3/>

²⁷Es sind derzeit oft einfach nur Metadaten dieses *Musikmetadaten*typs in den Musikdokumenten enthalten

angegeben sind, und dass sich die Schreibweisen der verschiedenen zugewiesenen Merkmale, wie z.B. Künstler und Musikstücktitel, nicht selten stark von einander unterscheiden. Im nächsten Abschnitt werden Ansätze vorgestellt, welche versuchen diesem Missstand zu beheben.

5.2.2 Metadatenanreicherung durch Informationssuche

Mit den durch den im vorherigen Abschnitt beschriebenen Metadatenextraktionsprozess gewonnenen *redaktionellen Metadaten* (ggf. auch Metadaten anderer *Musikmetadatentypen*) und den durch die Audiosignalextraktion und -anreicherung (siehe Abschnitt 5.1) vorhandenen *akustischen Metadaten* kann nun versucht werden das Profil für ein Musikstück, einen Künstler oder auch eine Musikzusammenstellung²⁸ vervollständigt und mit validierten Informationen, insbesondere auch *kulturellen Metadaten*, angereichert zu werden.

Zahlreiche Musikinformationsquellen im Internet bieten umfangreiche Metadatendienste an (siehe Abschnitt 4.3), welche zur Metadatenanreicherung genutzt werden können. Je geringer die schon vorhandenen *redaktionellen Eigenschaften* im Modell sind, umso mehr muss Verlass auf die extrahierten *akustischen Metadaten* sein. Sie sollen ggf. für eine Grundinitialisierung der bibliografischen Beschreibungen, d.h. eine Zuordnung von Werten für die *Metadatenfelder* Künstler, Musikstücktitel und Albumtitel²⁹ des Musikstückes (vgl. [Rai08a]), sorgen bzw. die vorhandenen Informationen mit validieren. Die Identifikation findet i.A. mit dem erstellten *musikalischem Fingerabdruck* des lokalen Musikstückes statt, welcher an einen *Musikdatenbankdienst* (z.B. *MusicBrainz* oder *CDDDB*) geschickt und analysiert wird. Bei erfolgreicher Zuordnung kann dieser Dienst dann weitere Informationen zu diesem Element bereitstellen.

Die nun vorhandene Informationsmenge kann direkt zur Anreicherung der *Musikmetadaten* genutzt werden, indem Anfragen

- an *Web Services* von Metadatendiensten, z.B. *Last.fm*, *Discogs* oder *Amazon* (d.h. besonders für das *Gemeinschaftliche Filter*,
- an allgemeine Suchmaschinen, z.B. *Google*, bzw. speziellen Musikinformationenseiten, z.B. *The Hype Machine*³⁰, mit anschließender *NLP*-Auswertung (d.h. besonders für das *kontextbasierte Filtern* siehe z.B. [SW08] und [KPS⁺08]),
- oder direkt an *SPARQL*³¹-Endpunkte, wie z.B. *DBPedia*³², *DBTune*³³ oder *BBC Backstage*³⁴,

gestellt werden. An den zuletzt aufgeführten *SPAQRL*-Endpunkten können direkt *SQL*-ähnlichen Anfragen mit *HTTP-URIs* auf Wissenbasen (auch Datensatz genannt; engl. *dataset*) mit *semantischen Graphen* gestellt werden. Das bedeutet insbesondere auch, dass

²⁸Diese Liste lässt sich auch noch beliebig fortsetzen, z.B. Plattenfirma, etc.

²⁹Wobei manche Songs auch gar nicht irgendeiner Musikzusammenstellung angehören, sondern lose existieren (besonders im unveröffentlichten Zustand)

³⁰Als Musik-Blog-Aggregator; siehe <http://hypem.com/>

³¹engl. *SPARQL Protocol and RDF Query Language*; siehe <http://www.w3.org/TR/rdf-sparql-query/> und <http://www.w3.org/TR/rdf-sparql-protocol/>

³²<http://dbpedia.org/sparql>

³³<http://dbtune.org/>

³⁴Welcher die *Music Ontology* mit nutzt, siehe [ian09], <http://bbc.openlinksw.com/sparql> und <http://api.talis.com/stores/bbc-backstage>

SPARQL Named Graphs verarbeiten kann und somit Anfragen mit einem bestimmten *semantischen Graphen* gemacht werden können, welcher durch eine *URI* repräsentiert wird (siehe [Rai08a]). Die Ergebnismenge ist dann wieder eine Beschreibung in Form eines *semantischen Graphens*. Diese Informationssuche kann dann ggf. beliebig fortgesetzt werden, bis ein gewünschter Wissenstand oder Grenzwert erreicht ist.

```

PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX tags: <http://www.holygoat.co.uk/owl/redwood/0.1/tags/>
SELECT ?album ?title ?date ?tag
WHERE
{
  ?album a mo:Record;
  foaf:maker <http://dbtune.org/jamendo/artist/1023>;
  ?album dc:title ?title;
  ?album dc:date ?date;
  ?album tags:taggedWithTag ?tag.
}
ORDER BY ?album

```

Auflistung 5.1: Eine SPARQL-Anfrage an den *Jamendo*-Datensatz von *DBTune*

In Auflistung 5.1 ist ein Beispiel einer *SPARQL*-Anfrage an den *SPARQL*-Endpunkt des *Jamendo*-Datensatzes³⁵ von *DBTune* zu sehen. In dieser Anfrage werden alle Musikalben (*mo:Record*) eines bestimmten Künstlers (*foaf:maker*; von dem schon eine zugehörige *URI* bekannt ist) mit ihrem Titel (*dc:title*), ihrem Veröffentlichungsdatum (*dc:date*) und zugehörigen Kennzeichnungen (*tags:taggedWithTag*) als *semantischer Graph* ausgegeben. Die *SPARQL*-Anfragesprache macht dabei Gebrauch von der kompakten und gut lesbaren *RDF*-Notation *Turtle/N3*, welche auch zur Ausgabe der Ergebnisse genutzt werden kann.

Verknüpfungstechniken und Ontologien, wie z.B. *void*³⁶ (engl. *Vocabulary of Interlinked Datasets*), *SIOC*³⁷ oder *MOAT*³⁸, helfen bei dieser rekursiven Suche im *Web of Data*³⁹ (*WoD*; auch *Linked Data* genannt) den Benutzer mit gehaltvollen Informationen zu versorgen (siehe [CH09], [PR08] und [KSB⁺08]). Eine wichtige Grundlage ist dabei das *Linking Open Data*-Projekt[mis09e] (*LOD*), welches bestrebt ist, semantisch verknüpfte Informationen im Internet verfügbar zu machen (z.B. die genannten *SPARQL*-Endpunkte), indem es auch bereits verfügbares Wissen ausnutzt, z.B. von *Wikipedia*, *Geonames* oder über in Webseiten eingebettete *Microformats*⁴⁰ oder *RDFa*-Beschreibungen (siehe [mis09e]).

Die *SPARQL*-Endpunkte nutzen dabei ggf. verschiedene Abbildungstechniken z.B. von relationalen Datenbanken (z.B. *D2R Server* mit *D2RQ*-Abbildungen⁴¹, oder von *Prolog*-

³⁵<http://dbtune.org/jamendo/#query>

³⁶<http://rdfs.org/ns/void/html>

³⁷*Semantically-Interlinked Online Communities*; siehe <http://sioc-project.org/>

³⁸*Meaning Of A Tag* <http://moat-project.org/>

³⁹Als Vision des *Semantic Web* definiert durch Tim Berner-Lee; siehe [BL09]

⁴⁰<http://microformats.org/>

⁴¹siehe <http://www4.wiwiw.fu-berlin.de/bizer/d2r-server/>

Prädikaten (z.B. *Prolog-2-RDF*⁴²) auf *semantische Graphen*. Neben der direkten Ausgabe von *semantischen Graphen* können z.B. Ausgabeformate (oft spezielle *XML*-Dialekte, z.B. *XHTML/Microformats*) von Webseiten, *Webservices* oder von (eventuell lokalen) *Audiosignalextraktionsanwendungen* mit Hilfe von *GRDDL*⁴³ und *XSLT*⁴⁴ in *semantische Graphen ad-hoc* umgewandelt und somit effizient weiter verarbeitet werden (siehe z.B. [Rai08b]).

Die Metadatenanreicherung kann lokal auf dem Computer bzw. innerhalb eines privaten Netzwerkes fortgesetzt werden. Dabei können insbesondere die Informationen aus Benutzerprofilen (vgl. [Her08]), aus vorhandenen Musikbibliotheken (z.B. *iTunes Music Library*, vgl. [mis07]) bzw. aus dem Umfeld des *Semantic Desktop* mit einfließen, wobei die Technik der *SPARQL*-Endpunkte hier gut zum Einsatz kommen kann (vgl. [Rai08a]). Somit können die Vorlieben und andere Statistiken der Anwender mit auf die Informationssuche Einfluss nehmen, aber auch Relation zu Dokumenten anderer Medientypen hergestellt werden.

Letztendlich soll nach diesem Schritt eine Wissensbasis über alle drei *Musikmetadatentypen*, d.h. *redaktionelle*, *kulturelle* und *akustische Metadaten*, existieren, auf der Strukturierungs- und Kategorisierungsalgorithmen angewandt werden können, welche im nächsten Abschnitt kurz erklärt werden.

5.2.3 Klassifikation, Zuordnung und Ähnlichkeiten von Musikmetadaten

Die Einordnung (Gruppierung), Klassifikation, Regression, Komprimierung (Daten-/ Dimensionsreduktion) und Ähnlichkeitsmessung von Musikmetadaten ist eine sehr komplexe Aufgabe nach der grundlegenden Merkmalsextraktion. Die Bereiche „Kurzzeitgedächtnis“ und „Langzeitgedächtnis“, des am Anfang dieses Kapitels vorgestellten *Musiksignal-Analyserahmenwerkes*, werden durch diesen Schritt abgedeckt (siehe Abbildung 5.1). Die Grundlagen dafür bilden die weitreichenden *multivariaten Analysemethoden*, worauf die Algorithmen des *Maschinellen Lernens* aufbauen. In den Büchern von Backhaus et al.[BPEW08], Camastra et al.[CV08] und Bishop[Bis06] sind die im Folgenden vorgestellten Methoden und Techniken sehr gut und ausführlich erklärt. Ein gutes Rahmenwerk, welches eine Vielzahl der erwähnten Algorithmen und Techniken implementiert, ist das *Java*-basierte *Weka*-Rahmenwerk[WF05].

5.2.3.1 Multivariate Analysemethoden

Die *multivariate Statistik* befasst sich mit der Auswertung der Zusammenhänge mehrerer Variablen bzw. Merkmale von Daten und findet Anwendung in den *multivariaten Analysemethoden* (vgl. [BLD⁺08]). Eine grobe, aber nicht überschneidungsfreie Zuordnung der Analysemethoden lässt sich nach ihrem Anwendungsbezug vornehmen (siehe [BPEW08]). Dadurch erfolgt eine primäre Einordnung entweder zu den *explorativen Verfahren*, d.h. strukturentdeckenden Verfahren, oder zu den *konfirmativen Verfahren*, d.h. strukturprüfende Verfahren. Im Folgenden werden nun wesentliche Verfahren⁴⁵ kurz zur

⁴²siehe <http://moustaki.org/p2r/>

⁴³*Gleaning Resource Descriptions from Dialects of Languages*, siehe [mis07]

⁴⁴*XSL Transformations*

⁴⁵Eine umfangreichere Liste findet man z.B. in [BPEW08]

Übersicht genannt und danach jeweils erläutert:

- **Explorative Verfahren:**
 - *Faktorenanalyse*
 - *Clusteranalyse*
 - *Multidimensionale Skalierung (MDS)*
 - *Künstliche Neuronale Netze*
- **Konfirmative Verfahren:**
 - Regressionsanalyse
 - Diskriminanzanalyse

Faktorenanalyse Die Faktorenanalyse hat es sich zum Ziel gesetzt, auf der einen Seite große Datenmengen durch Gruppierung von hochkorrelierenden Variablen zu strukturieren, und auf der anderen Seite durch Ermittlung von Ausprägungen der Struktur eine Datenreduktion vorzunehmen. Diese kompakten Faktoren⁴⁶ können dann anstelle der Originaldatenmenge genutzt werden.

Die Wahl des Faktorextraktionsverfahren hängt eng mit der Bestimmung der *Kommunalität* zusammen, d.h. wie viel Varianz kann der Faktor noch im Gegensatz zu seinen Ausgangsvariablen repräsentieren. Dabei sind die zwei wesentlichen Verfahren die *Hauptkomponentenanalyse* (engl. *Principal Component Analysis, PCA*) und die *Hauptachsenanalyse*, welche beide iterativ sind. Die erste Analyseverfahren unterstellt den Ausgangsvariablen dabei keine Einzelrestvarianz und ist deshalb eher zur Datenreduktion geeignet (siehe [BPEW08]), wie sie auch im *MIR* eingesetzt wird (z.B. bei der Bestimmung von *MFCCs*, siehe [Log00]). Die zweite Analyseverfahren hingegen betrachtet die Varianzen der Variablen und versucht damit Ursachen für eine hohe Korrelation zu klären. Im Einsatzbereich der Algorithmen des *Maschinellen Lernens* gilt die *Hauptkomponentenanalyse* als eines der am häufigsten angewendeten Verfahren zur Dimensionsreduktion und beschleunigt darüber hinaus deren Berechnung (vgl. [CV08] und [Jeh05]).

Clusteranalyse Die *Clusteranalyse* beschäftigt sich im Wesentlichen mit der Gruppenbildung von möglichst heterogenen, vorher oft unbekannt Gruppen auf möglichst homogenen Eigenschaften und Merkmalen, d.h. bei welchen die Objekte eine hohe Ähnlichkeit aufweisen. Dabei werden entweder Ähnlichkeiten (z.B. *Jaccard-Koeffizient*) oder Unähnlichkeiten (Distanzen; z.B. *Euklidische Distanz*) gemessen. Die darauf angewendeten Clusterverfahren lassen sich in zwei wesentliche Bereiche unterteilen. Auf der einen Seite sind dies die *partitionierenden Verfahren* und auf der anderen Seite die *hierarchischen Verfahren*.

Bei den *partitionierenden Verfahren* wird eine Vorgruppierung der Datenmenge mit einer festgelegten Anzahl vorgenommen und die Elemente so lange zwischen den einzelnen Gruppierungen hin- und hergetauscht, bis eine gegebene Zielfunktion ein Optimum erreicht (siehe [BPEW08]). Eine sehr bekannte Form sind dabei die *prototypbasierenden Clusterverfahren (PBC)*, welche es sich zum Ziel gesetzt haben, Repräsentanten für eine Gruppierung zu finden, denen die restlichen zugehörigen Objekte ähneln (siehe [CV08]).

⁴⁶Was den an früherer Stelle schon eingeführten *Merkmalsvektoren* entspricht

Die Grundlage für viele Clusterverfahren bildet dabei der sogenannte *EM-Algorithmus* (engl. *Expectation-Maximization-Algorithm*), welcher zur *Maximum-Likelihood-Schätzung* (engl. *Maximum-Likelihood-Estimation*, *MLE*) genutzt wird, d.h. wenn die Untersuchung nur auf Stichproben bzw. Auszügen aus einer großen Datenmenge vorgenommen werden kann. So findet dieser Algorithmus z.B. Anwendung in der *K-Means-Prozedur* oder in modellbasierten Ansätzen, wie z.B. *GMMs* oder *HMMs* (siehe [CV08, Bis06]), welche sehr häufig im *MIR* vorkommen (siehe z.B. [Tza02, Whi05]).

Im Gegensatz zu den *partitionierenden Verfahren* wird bei den *hierarchischen Verfahren* eine einmal gebildete Gruppierung nicht mehr aufgelöst [BPEW08]. Eine weitere Unterteilung findet bei dieser Kategorie durch die *agglomerativen* und *diversiven Algorithmen* statt. Die *agglomerativen Verfahren* gehen dabei von der feinsten Partition aus, d.h. jedes Objekt bildet eine eigene Gruppierung, und die *diversiven Verfahren* genau umgekehrt von der größten Partition, d.h. alle Objekte befinden sich in einer Gruppierung. Die hierarchische Struktur in Gruppierungen und Untergruppierungen usw., wie sie in diesem Bereich erzeugt werden, kann i.A. von *PBC-Methoden* nicht so effektiv gelöst werden (siehe [CV08]). Am verbreitetsten sind in dieser Kategorie die *agglomerativen Algorithmen* mit ihrem bekannten Vertreter, dem *Nearest-Neighbour-Verfahren* (*Nächstgelegener Nachbar*), welches die Objekte mit den kleinsten Distanzen zusammenfasst (vgl. [BPEW08]) und in verschiedensten Varianten im *MIR* angewendet wird (z.B. [SKW08]).

Im Allgemeinen bietet eine überschneidende Zuordnung mit Wahrscheinlichkeiten bzw. Gewichten mehr Freiheiten bei der Anpassung der Klassifikation an bestimmte Einflüsse, wie z.B. die die durch ein Benutzerprofil vorgenommen werden können (vgl. [CBF06]).

Multidimensionale Skalierung Die Methoden der *Multidimensionalen Skalierung* dienen im Wesentlichen zur Datenreduktion unter weitestgehenden Beibehaltung der Distanzen zwischen den Daten [CV08]. Dabei findet eine Darstellung von Eigenschaften in einem zwei- oder mehrdimensionalen Raum statt, um somit insbesondere die subjektive Wahrnehmung von Objekten durch Benutzer zu visualisieren [BPEW08]. Diese Abbildung ähnelt dabei in gewisser Weise auch der *Faktorenanalyse*, wobei die Positionierung mittels *MDS* auf der Beurteilung der Un-/ Ähnlichkeiten zwischen den Objekten basiert. Somit können die relevanten Eigenschaften durchaus unbekannt sein und es muss auch keine Auswahl dieser erfolgen (siehe [BPEW08]).

Eine Art Erweiterung stellt dabei die sogenannte *Isomap*⁴⁷ dar, welche die Unähnlichkeiten in Form von *geodätischen Distanzen*⁴⁸ in Bezug auf die (mathematische) *Mannigfaltigkeit* (z.B. eine Kugeloberfläche) ausdrückt (siehe [Bis06]). Somit eignet sich dieses Verfahren besonders für nichtlineare Dimensionsreduktion. Unter gewissen Voraussetzungen kann die *Isomap*-Methode theoretisch sogar die Mannigfaltigkeit unter einer sehr hohen Genauigkeit rekonstruieren (siehe [CV08]).

Eingesetzt wird die Technik der *MDS* bzw. *Isomap* z.B. zur Beschreibung der Klangfarbe in Wahrnehmungsräumen (siehe [BM07]) oder zur Beschreibung von Klangsegmenten (siehe [Jeh05]).

Künstliche Neuronale Netze Die Analysetechnik der *KNNs* orientiert sich an den Abläufen des menschlichen (und tierischen) Nervensystems und versucht diese durch mathematische Operationen nachzuvollziehen [BPEW08]. Es werden dabei i.A. keine

⁴⁷als Akronym für engl. *Isometric Feature Mapping*[CV08], d.h. *isometrische Merkmalsabbildung*

⁴⁸d.h. lokal kürzeste Verbindungskurve zweier Punkte

Vorgaben zwischen den Eingabeeigenschaften und den Ausgabemerkmale gemacht, d.h. keine kausalen Verknüpfungen oder die Annahme lineare Zusammenhänge. Diese Verfahren können klassische Verfahren der *multivariaten Analyse* generell in den Bereichen der Klassifikation (z.B. *Diskriminanzanalyse*), Prognose (z.B. *Regressionsanalyse*), Gruppenbildung (z.B. *Clusteranalyse*) oder Dimensionsreduktion (z.B. *Faktorenanalyse*) ersetzen, da sie recht gut hochgradig nicht-lineare und komplexe Zusammenhänge ohne spezifisches Vorwissen modellieren können (vgl. [BPEW08]).

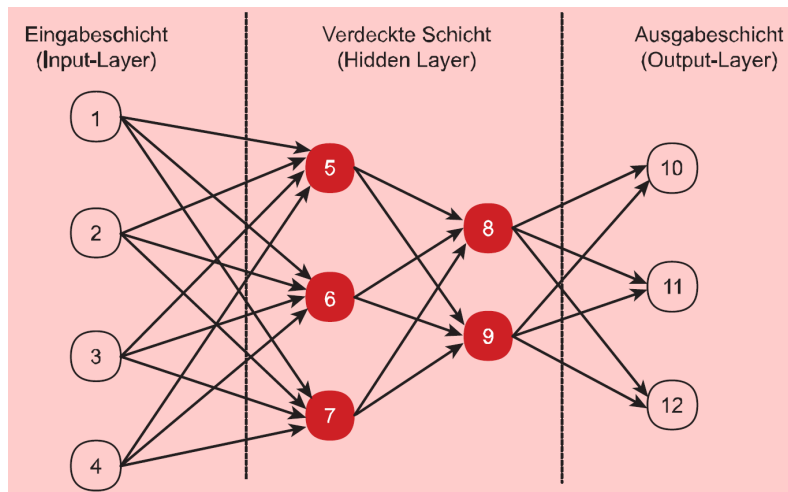


Abbildung 5.6: Grundstruktur eines (dreischichtigen) künstlichen neuronalen Netzes [BPEW08]

Anfänglich wird ein KNN, wie bei Algorithmen des *Maschinellen Lernens* (siehe nächster Abschnitt) üblich, mit beobachteten Daten trainiert, damit sich bestimmte Grundstrukturen ergeben und z.B. die Gewichte für verschiedene Beurteilungen angepasst werden um die Fehlerrate zu verringern. Die Informationsverarbeitung kann dabei vorwärts- oder rückwärtsgerichtet sein. Von großer allgemeiner Bedeutung, aber auch speziell im Bereich des *MIR* (siehe z.B. [Par07]), ist das sogenannte *Multi-Layer-Perceptron (MLP)*, welches vorwärts gerichtet ist und somit eine gute parallele Verarbeitung der Daten ermöglicht. Dieses *KNN* besteht aus den drei Grundschichten: Eingabeschicht, verdeckte Schicht und Ausgabeschicht, wobei die Anzahl der verdeckten Schichten tendenziell größer 1 sein kann (siehe [CV08]). In Abbildung 5.6 ist ein Beispiel eines *MLP* mit zwei verdeckten Schichten zu sehen, wobei die einzelnen Knoten als *Künstliche Neuronen* bezeichnet werden, welche die Eingabewerte verarbeiten und wichten (siehe [CV08]).

Ein weiteres populäres Beispiel sind die *Selbstorganisierenden Karten* (engl. *Self-organizing Map, SOM*) in all ihren Variationen, wobei i.A. ein hochdimensionierter Merkmalsraum auf eine zweidimensionale Karte abgebildet wird, so dass sich ähnliche Elemente nah aneinander befinden. Es wird z.B. zur automatischen Musikgenrebestimmung (siehe z.B. [Pam06]) oder zur automatischen Wiedergabelistengenerierung eingesetzt (siehe z.B. [DSPK08]).

Regressionsanalyse Die *Regressionsanalyse* versucht Beziehungen zwischen abhängigen, und ein oder mehreren unabhängigen Variablen (Eingabewerten) herzustellen, um darauf aufbauend Prognosen abzugeben, d.h. Vorhersagen über bestimmte Charakteristiken zu treffen, z.B. die Zugehörigkeit zu einer Gruppierung (vgl. [BPEW08] und

[CV08]). So kann z.B. mit der *Logistischen Regression* die Wahrscheinlichkeit der Zugehörigkeit zu einer Gruppe oder Kategorie in Abhängigkeit von einer oder mehrerer unabhängigen Variablen bestimmt werden (ähnlich wie bei der *Diskriminanzanalyse*, siehe nächster Paragraph). Das wesentliche Ziel bei der *Regressionsanalyse* besteht darin, eine *Regressionsfunktion* zu finden, bei der die nicht erklärbaren Abweichungen möglichst gering sind (siehe [BPEW08] und [Bis06]). Ein wichtiges Anwendungsgebiet der *Regressionsanalyse* sind die *Zeitreihenanalysen*, wobei es um die Beschreibung und Veränderung der zeitlichen Entwicklung einer Eigenschaft und deren Prognose geht (z.B. Grundschlagschätzung, siehe Abschnitt 5.1.5.1 oder im Speziellen Erkennung von langsamen Grundschlägen [Jeh05]).

Diskriminanzanalyse Die *Diskriminanzanalyse* ist ein Verfahren zur Analyse von Gruppenunterschieden [BPEW08], d.h. z.B. die Begründung der Unterscheidungsmerkmale zweier Musikgenre. Des Weiteren dient diese Technik im Wesentlichen zur Klassifizierung von Elementen. Sie definiert deshalb verschiedene Eigenschaften (als Parameter der *Diskriminanzfunktion*), welche sich besonders zur Unterscheidung von Gruppen der Problemstellung eignen, und nutzt die Merkmalsausprägungen der Elemente dann zur einer möglichen Gruppenzuweisung. Das Ziel der *Diskriminanzanalyse* ist die Minimierung der Wahrscheinlichkeit einer Fehlklassifikation (siehe [BPEW08]). In ihrem Zweck steht die *Diskriminanzanalyse* hinter der *Clusteranalyse*, wobei sich beide Verfahren wechselseitig ergänzen können: Die *Clusteranalyse* indem sie ihre entwickelte Taxonomie durch die *Diskriminanzanalyse* verifiziert, und die *Diskriminanzanalyse* indem sie Informationen, wie z.B. Fehlklassifikation, an die *Clusteranalyse* weiterleitet, welche darauf hin versucht ihre Taxonomie anzupassen. Als Maß der Unterschiedlichkeit (*Diskriminanzkriterium*) wird i.A. die Streuung zwischen den Gruppen mit der Streuung in den Gruppen jeweils auf Basis der Gruppenschwerpunkte betrachtet (siehe [BPEW08]). Ein Beispiel für eine *Diskriminanzfunktion* ist der *Bayes-Klassifikator*, welcher auch im *MIR* eingesetzt wird (siehe z.B. [Pfl08]).

5.2.3.2 Algorithmen des Maschinellen Lernens

Die im vorherigen Abschnitt beschriebenen *multivariaten Methoden* bilden die Grundlage für die Algorithmen des *Maschinellen Lernens* bzw. werden von diesen einfach oder in Kombination wiederverwendet (z.B. *KNN*). Das *Maschinelle Lernen* lässt sich als die Anwendung und Erforschung von Lernprozessen in ihrer vielfältigen Erscheinungsform definieren, welche mit Hilfe von berechenbaren Modellen Computersysteme dazu befähigen, selbstständig Wissen aufzunehmen und dieses zu erweitern, um ein gegebenes Problem danach besser lösen zu können (vgl. [CV08] und [BLD⁺08]). Dabei gibt es verschiedene Formen des Lernens, z.B. *mechanisches Lernen*, *Lernen durch Unterweisung* oder *Lernen durch Beispiele*, wobei die letzte Kategorie i.A. auf die im *MIR* angewandten Techniken des *Maschinellen Lernens* zutrifft.

Diese Methode ermöglicht es einem *Lernenden*, d.h. einer *Lernmaschine*, aus einer gegebenen Menge an Beispielen eines Konzeptes eine verallgemeinernde Konzeptbeschreibung abzuleiten. Der Programmierer einer solchen *Lernmaschine* kann deshalb auch als *Lehrer* dieser angesehen werden, da er das Wissen besitzt, um die Problemlösung umzusetzen (vgl. [CV08]). Aus dieser Beschreibung heraus lassen sich auch die drei wesentlichen Kategorien zur Unterteilung des *Maschinellen Lernens* ableiten. Sie heißen

überwachtes Lernen, *bestärkendes Lernen* und *unüberwachtes Lernen*, wobei auch noch Mischformen (siehe [Bis06] und [CV08]), wie z.B. *halbüberwachtes Lernen*, existieren (siehe [Zhu08]). Wo beim *überwachten Lernen* der *Trainingsdatensatz*, d.h. die Beispiele zum Erlernen des Konzeptes, aus Eingabedaten (im Speziellen die verschiedenen Merkmalsvektoren) und Ausgabedaten (die vom *Lehrer* festgelegten Kategorien) besteht, bekommt die *Lernmaschine* beim *unüberwachten Lernen* nur die Eingabewerte. Dem entsprechend benutzt das *halbüberwachte Lernen* gekennzeichnete, d.h. mit Beschreibung der Ausgabedaten, und nicht gekennzeichnete Daten als *Trainingsdatensatz*. Algorithmen, welche auf dem Prinzip des *Bestärkendes Lernen* beruhen, z.B. *HMMs*, versuchen mit Belohnungen und Bestrafungen das Verhalten der *Lernmaschine* zu optimieren, denn sie weiß nicht die optimale Handlung zu einem bestimmten Zustand wie beim *überwachten Lernen* (siehe [CV08]).

Die Methoden der *Clusteranalyse*, der *Faktorenanalyse* und der *MDS* befinden sich sehr oft im Bereich des *unüberwachten Lernens*. Dahingegen werden die Techniken der *Regressions-* und *Diskriminanzanalyse* sehr oft mit in den Bereich des *Überwachten Lernens* einbezogen, z.B. in den populären und leistungsfähigen *Support-Vektor-Maschinen (SVM)*; siehe z.B. [Whi05] oder in *Bayes'schen Netzen* (siehe z.B. [RP08]). Letztendlich finden die Methoden des *KNN* in allen Bereichen Anwendung, z.B. *SOM* beim *unüberwachten Lernen* und *MLP* beim *Überwachten Lernen* (vgl. [CV08]).

Abschließend sind auch noch die Kombinationen der verschiedenen Modelle des *Maschinellen Lernens* zu erwähnen, welche die Ergebnisse der einzelnen *Lernmaschinen* zu einer Gesamtheit zusammenfügen. Im *MIR* finden sich dabei u.a. die Technik der *Entscheidungsbäume* (siehe z.B. [Pfl08]), der *Mixture-of-Experts* (siehe z.B. [Par07]), der *GMMs* oder des *Boosting* (siehe z.B. [BMEML08]). Besonders bei diesen Methoden kann von der Technik *Dynamisches Programmieren*⁴⁹ Gebrauch gemacht werden, welche versucht dynamische Optimierungsprobleme auf Basis der Rekursion zu lösen, z.B. auf Zeitstrukturen (siehe [Jeh05]).

5.2.3.3 Ontologien zur Klassifikation und Ähnlichkeitsanalyse

Ontologien können auch für den Bereich der Klassifikation und Ähnlichkeitsanalyse als eine hochwertiges Datenmodell verwendet werden. So können die durch die Klassifikation und Zuordnung gefundenen Kategorien und Gruppierungen z.B. mit Hilfe der *SKOS-Ontologie* beschrieben werden. Diese Ontologie ist speziell für die vier *Wissensorganisationssysteme* Glossar, Taxonomie, Thesaurus und Klassifikationsschema von der *SWD-WG*⁵⁰ entworfen worden und wird von dieser auch als *Best Practice* empfohlen (siehe [MMWB05]). Sie ermöglicht somit nicht nur eine hierarchische Beschreibung in Form einer Taxonomie (siehe Abschnitt 3.2), sondern kann auch diverse Zusammenhänge (z.B. Querverweise) zwischen den verschiedenen Anhäufungen von Objekten mit ähnlichen Eigenschaften beschreiben. Zum Beispiel wurde sie u.a. zur Definition einer Instrumenten-Taxonomie als eine Erweiterung der *Music Ontology* genutzt (siehe Abschnitt 4.2.2.2 Paragraph „Music Ontology“).

Daneben lassen sich durch die Anwendung der *Similarity-Ontologie*⁵¹ auf der Ebene von Ontologien diverse und ggf. auch komplexe Ähnlichkeitsanalysen von Musikmetadaten

⁴⁹Als mathematische Regelmenge zur Lösung eines Problems [Jeh05]

⁵⁰W3C Semantic Web Deployment Working Group, siehe <http://www.w3.org/2006/07/SWD/>

⁵¹<http://purl.org/ontology/similarity/>

auf Basis der *Music Ontology* (inkl. ihrer Erweiterungen) durchführen. Mit dieser Ontologie können z.B. nicht nur Zusammenhänge zwischen den Eigenschaften aus der Merkmalsextraktion beschrieben werden, sondern auch subjektive Zusammenhänge, die durch einen Benutzers definiert werden, z.B. ein bestimmtes Erlebnis und die damit in Erinnerung gebliebenen Musikstück oder Künstler. Jede Assoziation, d.h. z.B. eine komplexe Beschreibung von Einflüssen, kann dabei über eine separat definierte Assoziationsmethode (*AssociationMethod*) definiert werden, z.B. mit Hilfe von *Named Graphs* (siehe [JR09]).

5.2.4 Persönliche Musikwissenbasis

Nachdem die verschiedenen Musikmetadaten so weit wie möglich gesammelt und analysiert worden sind, sollte sie noch in einer gut wiederverwendbaren, persistenten und effizienten Form hinterlegt werden, damit u.a. auch eine Anwendung zur Verwaltung von persönlichen Musiksammlungen diese nutzen kann. Damit kann diese Informationsmenge die *persönliche Musikwissenbasis* über die eigene Musiksammlung bilden. Die in Kapitel 4 vorgestellten *Musikmetadatenformate* stellen dafür die Grundlage bereit. Dabei hat sich eine auf *semantischen Graphen* basierte Lösung auf Basis der *Music Ontology* als die beste Wahl herausgestellt, da sie so einfach wie möglich die semantischen Zusammenhänge von Musik und musikrelevanten Eigenschaften modellieren kann, auf denen die Entscheidungen zur Auswahl und Darstellung von Musikmerkmalen und -metadaten getroffen werden können.

Insbesondere ist auch nicht die Verwendung von Tripel, sondern die Verwendung von Quadrupel zum Abspeichern die bessere Wahl, um die komplexe Informationsmenge aus der Merkmalsextraktion effektiv zu verarbeiten (vgl. [Rai08a]). Somit können gut *Named Graphs* eingebunden werden (wie z.B. in *N3*[Rai08a]) und das Tripel aus Subjekt, Prädikat und Objekt erweitert sich zu einem Quadrupel mit den Elementen des Tripels und einem zusätzlichen, welches den Kontext des Tripels definiert. Der Kontextbezeichner kann so einfach wieder in anderen Tripel wiederverwendet werden (siehe [Rai08a]). Da viele, zur Zeit existierende Musikinformationsquellen, insbesondere die Anwendungen zur Audiosignalextraktion, oft noch keine auf *semantischen Graphen* basierten Ausgaben liefern, müssen diese ggf. mit den z.B. in Abschnitt 5.2.2 erwähnten Techniken umgewandelt werden. Eine Ausnahme bildet hier z.B. der *Sonic Annotator*⁵², welcher mit Hilfe von durch *VAMP*-Plug-ins (siehe Einleitung Abschnitt 5.1) berechnete und in *semantische Graphen* umgewandelte, Musikmerkmalsbeschreibungen auf Basis der *Music Ontology* und *Audio Feature*-Ontologie liefert (siehe Abschnitt 4.2.2.2 und [KC08]). Für einen schnellen Zugriff auf entstandene Merkmalsvektoren, z.B. zur erweiterten Ähnlichkeitssuche, empfiehlt sich die Verwendung von speziell auf die Verarbeitung dieser Konstrukte optimierten Datenbanksystemen, z.B. *audioDB*⁵³ (siehe [mis09a] und [MCR08] für die darauf aufbauende Technik *mHashup*⁵⁴).

Auf der anderen Seite sind die Technologien im Bereich der Verarbeitung und persistenten Abspeicherung von *semantischen Graphen* mittlerweile sehr leistungsfähig geworden. Zum Beispiel hat die Firma *OpenLink Software*⁵⁵ mit ihrer Lösung *Virtuoso Universal*

⁵²<http://omras2.org/SonicAnnotator>

⁵³<http://omras2.doc.gold.ac.uk/software/audiodb/>

⁵⁴<http://www.mhashup.com/>

⁵⁵<http://www.openlinksw.com/>

*Server*⁵⁶ ein hybrides Hochleistungserversystem entwickelt, welches u.a. eine effiziente Form der Hinterlegung von *semantischen Graphen* in einer Datenbank (*Triple Store* bzw. *Quad Store*) mit direkter Unterstützung der Anfragesprache *SPARQL*, den Zugang über einen *SPARQL*-Endpunkt und auch die Anbindung an Verarbeitungsrahmenwerke für *semantische Graphen* wie z.B. *Jena*⁵⁷ oder *Sesame*⁵⁸ bietet (siehe [Ope09a]). Glücklicherweise ist auch eine frei verfügbare Variante dieses Serversystems erhältlich (siehe [Ope09b]).

Eine Anbindung einer *audioDB*-Instanz, für die Merkmalsvektoren der Audiosignalextraktion und -anreicherung, an eine *Virtuoso*-Instanz, mit den Informationen aus der Metadatenanalyse und -anreicherung, wäre hinsichtlich einer hochperformanten Datenverarbeitung eine interessante Lösung, da die beiden System schon einzeln recht leistungsfähig in Bezug auf ihren Anwendungsbereich sind. Die *SPARQL*-Schnittstelle der *audioDB* eignet sich dabei ideal für delegierte Anfragen und liefert die Antwort in Form eines *RDF*-Graphen zurück (siehe [mis09b]).

Mit dieser persönlichen Musikwissenbasis, welche ggf. mit zusätzliche Informationen bei Bedarf erweitert werden kann (siehe Abschnitt 5.2.2), können die in Kapitel 2 vorgestellten, wesentlichen Anwendungsfälle, welche bei der Verwaltung einer privaten Musiksammlung vorkommen können, bearbeitet werden.

5.3 Fazit

Diese Kapitel zeigte die Komplexität des Prozesses der Merkmalsextraktion von Musik und musikrelevanten Eigenschaften auf, welcher notwendig ist, um eine umfangreiche *Musikwissenbasis* zu erstellen, welche zur Realisierung von vielfältigen Zugangsmöglichkeiten auf eine persönliche Musiksammlung benötigt wird. Auf der einen Seite wurden grundlegende, oft angewandte und gut geeignete Methoden, Techniken und Modelle zur inhaltlichen Analyse von Musiksignalen beschrieben, und die verschiedenen Teilbereiche, aus denen sich die Merkmalsvektoren zur Beschreibung eines Musiksignals zusammensetzten, vorgestellt und eingeordnet (in Bezug auf die Musikmerkmale).

Auf der anderen Seite wurde der umfangreiche Prozess der Metadatenextraktion und -anreicherung mit dafür relevanten und eingesetzten Vorgehensweisen vorgestellt. Weiterhin wurde auch die Bedeutung des Einsatzes von Algorithmen des *Maschinellen Lernens* hervorgehoben, da die extrahierten Merkmalsvektoren bzw. Metadaten noch einen komplexen Prozess der Klassifikation, Zuordnung, Regression und Ähnlichkeitsauswertung durchlaufen müssen, um in einer für den Endanwender geeigneten Repräsentationsform dargestellt bzw. zu dieser wieder aufbereitet werden zu können. Dabei hat sich herausgestellt, dass bestimmte Methoden und Techniken besser zur Bearbeitung bestimmter Teilbereiche sind als andere und deshalb auch mehr in diesen Anwendungsgebieten eingesetzt werden (siehe die jeweils aufgelisteten speziellen Techniken der vorgestellten Methoden in Abschnitt 5.2.3). Dies zeigt insbesondere, dass die Modellierung der nicht trivialen Abläufe bei der Wahrnehmung und Verarbeitung von Musiksignalen durch das menschliche Hörsystem und Gehirn immer noch eine große Herausforderung ist, welche zwar schon recht umfangreich erforscht wurde, dabei aber immer noch Verbesserungen bei den

⁵⁶<http://virtuoso.openlinksw.com/overview/index.htm>

⁵⁷<http://jena.sourceforge.net/>

⁵⁸<http://sourceforge.net/projects/sesame/files/Sesame%20/>

Details, z.B. den verwendeten Parametern, und der Komposition der Verfahren gemacht werden können.

Eine Optimierung der dazu genutzten Datenmodelle spielt dabei eine nicht unwesentliche Rolle. Es wird sich in der Zukunft noch zeigen, ob die vorgestellten, modernen und vielversprechenden, ontologiebasierten Ansätze und deren Wissensverknüpfung sich bei der Wissensmodellierung durchsetzen und auch den Prozess der Merkmalsextraktion von Musik und musikrelevanten Eigenschaften begleiten werden – bis hin zur Abspeicherung in einer (persönlichen) *Musikwissensbasis*.

6 Zusammenfassung und Ausblick

In diesem Beleg wurde sich umfassend mit den Musikmerkmalen und -metadaten im Anwendungsbereich von persönlichen digitalen Musiksammlungen auseinandergesetzt, um Wissen und Erkenntnisse über Anforderungen zu erlangen, die an ein *Musikmetadatenformat* gestellt werden, welches für eine *persönliche Musikwissensbasis* geeignet ist.

Dafür wurden zuerst die verschiedenen wesentlichen Anwendungsfälle untersucht, welche im privaten Umfeld auftauchen (siehe Kapitel 2). Benutzer einer Anwendung zur Verwaltung von Musiksammlungen, stellen i.A. zwar unterschiedliche Ansprüche an eine solche, doch ist generell davon auszugehen, dass die derzeitigen Möglichkeiten, ihren Bedürfnissen nicht gerecht werden. Es gibt schon viele interessante Zugangsmöglichkeiten, wie man eine Musiksammlung erleben und bearbeiten kann, doch sind viel davon entweder noch Bestand der Forschung (z.B. *Musikkomentierungsspiele*), werden in anderen Anwendungsbereichen eingesetzt (z.B. *Online-Musikempfehlung* und *-Entdeckung*) oder sind noch sehr neu (z.B. *mufin vision* und ähnliche Techniken), so dass diese noch nicht genügend bekannt sind. Die derzeitige Unterstützung der vorhandenen Programme für verschiedene Anwendungsfälle ist auch eher schlecht, denn allgemein gilt, dass ein Anwender auf seine Musiksammlung so einfach und intuitiv wie möglich zugreifen möchte, um diese zu bearbeiten und zu nutzen. Im Speziellen bedeutet dies z.B. für den Anwendungsfall *Suchen*, dass unnatürliche schlagwortbasierte textuelle Suchanfragen fast immer möglich sind, aber *Anfrage per natürliche Sprache* oder *Anfragen mit Audioinhalt* in diesem Bereich nicht zur Verfügung stehen. Des Weiteren fehlt es an intuitiven Darstellungen der persönlichen Musiksammlung (oder Teile dieser), denn die textuelle Repräsentation von Musikcharakteristiken und -eigenschaften ist immer noch sehr dominant. Zwar gibt es schon recht gute Realisierungen zur Anzeige von Coverbildern (z.B. *Cover Flow*), doch lassen noch weitere visuelle Sichten in Anwendungen zur Verwaltung von Musiksammlungen auf sich warten. Letztendlich müssen auch Mechanismen entwickelt werden, welche wirkungsvoll gegen das zu rasche Anwachsen von persönlichen Musiksammlungen zum Einsatz kommen können (z.B. *digitales Ausleihen*).

Mit dem Wissen und der Erkenntnis, dass ein Anwender seine Musiksammlung auf eine vielfältige Art und Weise benutzen und erfahren möchte und, dass dafür eine hohe Menge an Musikmerkmale und -metadaten notwendig sind, wurden die Themen der Kapitel 3 und 4 dieses Aufsatzes bearbeitet. Bei der Untersuchung der Musikmerkmale wurden zuerst die einfachen und danach die angereicherten Audiosignaleigenschaften vorgestellt und untersucht. Eine Betrachtung des Wahrnehmungs- und Verarbeitungsprozesses von Audiosignalen durch den Menschen bildete dafür die Grundlage. So zeigte sich, dass neben objektiv meßbaren physikalischen Eigenschaften des Audiosignals, auch noch eine Vielzahl von subjektiv wahrgenommen Hörsinnesempfindungen existieren, welche durch die Wissenschaftsbereiche der Psychoakustik und Musikpsychologie erforscht wurden und werden. Dabei stellte sich heraus, dass besonders diese Charakteristiken bei der Einordnung und Klassifikation von Musikstücken eine bedeutende Rolle spielen. Weiterhin können die, durch die Musiktheorie und -wissenschaft definierten, angereicherten Au-

diosignaleigenschaften, die Grundlage für abstraktere und ungenauere Beschreibungen bilden bzw. diese davon abgeleitet werden. Deshalb ergab auch die Untersuchung von Taxonomien in der Musik die Einsicht, dass es keine optimalen Definitionen für weitestgehend subjektive Gruppierungen, wie z.B. Musikgenre oder empfundene Stimmungen, gibt und aus diesem Grund Annäherungen an solche Klassifikationen unter dem Einfluss von verschiedenen Informationsquellen (Expertenwissen, *Folksonomien* aber auch inhaltlichen Analysen) realisiert werden müssen – unter Einbeziehung des eigenen Benutzerprofils, damit die Taxonomien an die persönlichen Sichtweisen angepasst werden können. Nach den eigentlichen Musikmerkmalen, wurden anschließend in Kapitel 4 Musik- und musikrelevante Metadaten betrachtet. Dafür wurde zuerst eine allgemeine Einführung über (*Musik-*)*Metadattentypen*, Anwendungsbereiche von *Musikmetadaten* und *Metadatenfeldergruppierungen* gegeben, um anschließend eine Reihe von *Musikmetadatenstandards* und *-onologien* auf diese Charakterisierungen zu untersuchen. Dabei wurde eine Unterteilung in *formatgebundene* und *formatunabhängige Spezifikationen* zur Beschreibung von Musikedokumenten und deren Inhalten vorgenommen, bei welcher sich klar die Nachteile der vorhandenen *formatgebundenen Spezifikationen* abzeichneten. Deshalb wurde auch die Ersetzung der Metadatenbeschreibungen innerhalb von Musikedokumenten durch eine *formatunabhängige Spezifikation* vorgeschlagen, welche auf einem ontologiebasierten Datenmodell beruht, das einfach und effizient semantische Zusammenhänge von Musikinformationen modellieren kann (z.B. die *Music Ontology*). Weiterhin zeigte die Analyse von *Musikmetadaten Diensten* einen Bedarf an einer einheitlichen ontologiebasierten Metadatenpezifikation für Musikmerkmale und -metadaten, denn diese Dienste können bei dem Prozess der Merkmalsextraktion sehr hilfreich sein, wenn Lösungen existieren, die das Wissen der verschiedenen Musikinformationsquellen sinnvoll zusammenführen, so dass es anschließend dem Anwender nach seinen Wünschen repräsentiert werden kann.

Der in den ersten drei Kapiteln (ohne Einführung) erarbeitete Kenntnisstand, wurde in dem letzten Kapitel dieser Belegarbeit (ohne Zusammenfassung) angewandt, in welchem eine allgemeine Betrachtung des komplexen Prozesses der *Musikmerkmalsextraktion* vorgenommen wurde. Das Ziel dieses Vorganges ist, eine (persönliche) *Musikwissenbasis* zu erstellen, welche als Grundlage für den Umgang mit persönlichen digitalen Musiksammlungen dient. Eine Zerlegung in zwei wesentliche Bereiche fand bei der Untersuchung der *Musikmerkmalsextraktion* statt. So wurde zuerst der Teilprozess der Audiosignalextraktion und -anreicherung vorgestellt und bewertet, und anschließend die darauf aufbauende Metadatenextraktion und -anreicherung. Es zeigte sich u.a., dass eine Vielzahl von Methoden zur *Zeit-Frequenz-Transformation* von Audiosignalen existieren, aber nicht jede für diese Aufgabe geeignet ist. Dennoch existieren Methoden, wie z.B. *VRT*, die besonders für den Vorgang der *Zeit-Frequenz-Transformation* von Musiksignalen verwendet werden können. Bei der Metadatenextraktion wurde deutlich, dass eine Vielzahl von unterschiedlichen Informationsquellen nötig sind, um die aus den Musikedokumenten gewonnenen Metadaten zu überprüfen und zu ergänzen. Doch wurden dafür auch effiziente Techniken, wie z.B. *SPARQL-Endpunkte*, vorgestellt, über die einfach Wissen angereichert werden kann. Weiterhin ist es auch wichtig, dass die extrahierten Merkmalsvektoren der *Musiksignalanalyse* und die angereicherten Metadaten dazu in einem komplexen Prozess der Einordnung, Klassifikation und Regression einbezogen werden, welcher von Algorithmen des *Maschinellen Lernens* Gebrauch macht. Abschließend wurden noch verschiedene Datenbanksysteme und -techniken vorgestellt, welche verdeutlichen, dass die Daten

der *Musikwissenbasis* in einer gut wiederverwendbaren, persistenten und effizienten Form abgespeichert werden können, z.B. Quadrupel von *semantischen Graphen*.

Ausblick Die in dieser Arbeit gewonnen Erkenntnisse können sehr gut in die Realisierung einer *Musikmerkmalsextraktionskomponente* einfließen, welche Bestandteil einer *Semantic Desktop*-Umgebung sein kann. Dabei ist es wichtig, das empfohlene *Musikmetadatenformat*, die *Music Ontology*, noch für verschiedene Anwendungsfälle zu erweitern, z.B. zum vollständigen Modellieren von Wiedergabelisten oder ein erweitertes Veröffentlichungskonzept, welches z.B. die Informationen des *Discogs*-Dienstes gut abbilden kann. Auch bei der *Musiksignalanalyse* können die verschiedene Extraktionsabläufe optimiert und neuwertige Techniken, wie z.B. *VRT*, eingesetzt werden. Darüber hinaus müssen die Einordnungs-, Klassifikations- und Regressionsvorgänge noch weiter verbessert bzw. gute Ansätze, wie z.B. *Boosting*, auch in reale Anwendungsszenarien getestet werden. Die Verwendung von *SPARQL*-Endpunkten und die damit geschaffene Verbindung zum *Web of Data* ermöglichen einer *Musikmerkmalsextraktionskomponente* in eine neue Dimension der Datenanreicherung vorzudringen, das sie sich „nur noch“ bedienen muss. Nicht zuletzt ist es aber auch wichtig, dass eine solche Lösung das *Web of Data* mit Informationen versorgt, welche es noch nicht hat, damit sich der *Informationsversorgungskreislauf* schließt und die *Informationsrückgewinnung* effektiv ablaufen kann. Mit einer solchen Wissensbasis im Hintergrund, kann eine *persönliche Musikwissenbasis* getrost die Herausforderung annehmen, den Benutzer einer Anwendung zur Verwaltung von Musiksammlungen mit Informationen zu versorgen, welchen ihm einen einfachen und intuitiven Umgang mit seiner persönlichen Musikkollektion ermöglichen.

Literaturverzeichnis

- [AAG06] Xavier Amatriain, Pau Arumi, and David Garcia. CLAM: A Framework for Efficient and Rapid Development of Cross-platform Audio Applications. In *Proceedings of ACM Multimedia 2006*, pages 951 – 954, Santa Barbara, California, USA, Oktober 2006.
- [Ama09a] Amazon.com. Advantage for Music [online]. 2009. Available from: <https://www.amazon.com/gp/seller-account/mm-product-page.html?topic=200329710> [cited 07.09.2009].
- [Ama09b] Amazon.com. Amazon.com Product Advertising API License Agreement [online]. 2009. Available from: https://affiliate-program.amazon.com/gp/advertising/api/detail/agreement.html?ie=UTF8&pf_rd_t=501&ref_=amb_link_83995711_12&pf_rd_m=ATVPDKIKX0DER&pf_rd_p=&pf_rd_s=assoc-center-1&pf_rd_r=&pf_rd_i=assoc-api-detail-4-v2 [cited 07.09.2009].
- [BDA⁺05] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. A Tutorial on Onset Detection in Music Signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, September 2005.
- [Bil93] J. A. Bilmes. Timing is of essence. Master’s thesis, Massachusetts Institute of Technology, 1993.
- [Bis06] Christopher M. Bishop. *Pattern Recognition and Maschine Learning*. Springer Science+Business Media, LLC, 2006.
- [BL09] Tim Berners-Lee. Linked Data [online]. Juni 2009. Available from: <http://www.w3.org/DesignIssues/LinkedData.html> [cited 20.08.2009].
- [BLCK⁺08] Tim Berners-Lee, Dan Connolly, Lalana Kagal, Yosi Scharf, and Jim Hendler. N3Logic : A logical framework for the world wide web. *Theory and Practice of Logic Programming*, 8(3):249 – 269, 2008.
- [BLD⁺08] Dr. Eggert Winter Dr. Dr. Jörg Berwanger Joachim Becker, Meinolf Lombino, Birgitta Dennerlein, Olaf Fischer, Beate Rabe, and Inge Kachel-Moosdorf. Gabler Wirtschaftslexikon Online [online]. Juni 2008. Available from: <http://wirtschaftslexikon.gabler.de/> [cited 25.08.2009].

- [BM07] John Ashley Burgoyne and Stephen McAdams. Non-linear scaling techniques for uncovering the perceptual dimensions of timbre. In *Proceedings of the International Computer Music Conference*, volume 1, pages 73 – 76, Kopenhagen, Dänemark, 2007.
- [BMA09] BMAT. Vericast. Technical report, Barcelona Music & Audio Technologies, 2009.
- [BMEML08] Thierry Bertin-Mahieux, Douglas Eck, Francois Maillet, and Paul Lamere. Autotagger: A Model For Predicting Social Tags from Acoustic Features on Large Music Databases. *Journal of New Music Research*, 37:115–135, 2008.
- [BPEW08] Klaus Backhaus, Wulff Plinke, Bernd Erichson, and Rolf Weiber. *Multivariate Analyseverfahren*. Springer Berlin / Heidelberg, 2008.
- [Bro91] Judith C. Brown. Calculation of a Constant Q Spectral Transform. *Journal of the Acoustical Society of America*, 89(1):425–434, 1991.
- [BVL⁺04] Wouter De Baene, André Vandierendonck, Marc Lemanc, Andreas Widmann, and Mari Tervaniemi. Roughness perception in sounds: behavioral and ERP evidence. *Biological Psychology*, 67:319–330, 2004.
- [BW01] Mark A. Bartsch and Gregory H. Wakefield. To Catch a Chorus: Using Chromabased Representations for Audio Thumbnailing. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 15–18, Mohonk, New York, USA, Oktober 2001.
- [Can06] Pedro Cano. *Content-based Audio Search: From Fingerprinting to Semantic Audio Retrieval*. PhD thesis, Technology Department of Pompeu Fabra University, Barcelona, 2006.
- [CBF06] Sally Jo Cunningham, David Bainbridge, and Annette Falconer. 'More of an Art than a Science': Supporting the Creation of Playlists and Mixes. In *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Kanada, Oktober 2006.
- [CDH⁺07] Oscar Celma, Stamatia Dasiopoulou, Michael Hausenblas, Suzanne Little, Chrisa Tsinaraki, and Raphaël Troncy. MPEG-7 and the Semantic Web [online]. August 2007. Available from: <http://www.w3.org/2005/Incubator/mmsem/XGR-mpeg7-20070814/> [cited 04.02.2009].
- [CGD06] Nik Corthaut, Sten Govaerts, and Erik Duval. Moody Tunes: The Rockanango Project. In *Proceedings of the 7th International Conference of Music Information Retrieval*, Victoria, Kanada, Oktober 2006.
- [CGVD08] Nik Corthaut, Sten Govaerts, Katrien Verbert, and Erik Duval. Connecting the Dots: Music Metadata Generation, Schemas and Applications. In *Proceedings of the Ninth International Conference on Music Information Retrieval*, Philadelphia, USA, September 2008.

- [CH09] Richard Cyganiak and Michael Hausenblas. *void Guide - Using the Vocabulary of Interlinked Datasets* [online]. Januar 2009. Available from: <http://rdfs.org/ns/void-guide> [cited 20.08.2009].
- [CIR⁺08] Gal Chechik, Eugene Ie, Martin Rehn, Samy Bengio, and Dick Lyon and. Large-Scale Content-Based Audio Retrieval from Text Queries. In *Proceeding of the 1st ACM international conference on Multimedia information retrieval*, Vancouver, British Columbia, Kanada, 2008.
- [CJJ04] Sally Jo Cunningham, Matt Jones, and Steve Jones. Organizing Digital Music for Use: An Examination of Personal Music Collections. In *Proceedings of the 5th International Conference on Music Information Retrieval*, Barcelona, Spanien, Oktober 2004.
- [Cli07] Stephanie Clifford. Pandora's Long Strange Trip [online]. Oktober 2007. Available from: <http://www.inc.com/magazine/20071001/pandoras-long-strange-trip.html> [cited 06.09.2009].
- [CLP⁺07] Bin Cui, Ling Liu, Calton Pu, Jialie Shen, and Kian-Lee Tan. QueST: Querying Music Databases by Acoustic and Textual Features. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 1055–1064, New York, NY, USA, 2007. ACM. doi:<http://doi.acm.org/10.1145/1291233.1291465>.
- [CV08] Francesco Camastra and Alessandro Vinciarelli. *Maschine Learning for Audio, Image and Video Analysis*. Springer-Verlag London Limited, 2008.
- [Del01] I. Deliege. Prototype Effects in Music Listening: An Empirical Approach to the Notion of Imprint. *Music Perception*, 18, no. 3:371–407, 2001.
- [DHSW02] Erik Duval, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel. Metadata Principles and Practicalities. *D-Lib Magazine*, 8(4):–, April 2002.
- [Dis09] Discogs. Quick Start Guide [online]. 2009. Available from: <http://www.discogs.com/help/quick-start-guide.html> [cited 12.02.2009].
- [dIMHR05] Helga de la Motte-Haber and Günther Rötter. *Musikpsychologie*. Laaber-Verlag, 2005.
- [Dow03] J. Stephen Downie. *Annual Review of Information Science and Technology 37*, chapter Music Information Retrieval (chapter 7), pages 295–340. Information Today, 2003.
- [DP06] Francois Deliege and Torben Bach Pedersen. Music Warehouses: Challenges for the Next Generation of Music Search Engines. In *Proceeding of 1st Workshop on Learning the Semantics of Audio Signals*, pages 95–105, Athen, Griechenland, Dezember 2006.
- [DSPK08] Markus Dopler, Markus Schedl, Tim Pohle, and Peter Knees. Accessing Music Collections Via Representative Cluster Prototypes in a Hierarchical

- Organisation Scheme. In *Proceedings of the 9th International Conference on Music Information Retrieval*, Philadelphia, USA, September 2008.
- [DWW⁺75] Richard P. Delone, Gary E. Wittlich, Allen Winold, Mary H. Wennerstrom, and Vernon L. Kliwer. *Aspects of Twentieth-Century Music*. Prentice Hall, 1975.
- [Eus09] D. Eustace. Loudness [online]. 2009. Available from: http://www.acoustics.salford.ac.uk/res/cox/sound_quality/index.php?content=loudness [cited 23.05.2009].
- [Fin04] Michael Fingerhut. Music Information Retrieval, or how to search for (and maybe find) music and do away with incipits. In *International Association of Sound and Audiovisual Archives (IASA) Conference 2004*, Oslo, Norwegen, 2004.
- [Fra09] Fraunhofer IDMT. AudioID [online]. Januar 2009. Available from: http://www.idmt.fraunhofer.de/eng/research_topics/audioid.htm [cited 19.08.2009].
- [FZ07] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics: facts and models*. Springer, 3. edition, 2007.
- [Gal91] Antony Galton. Reified Temporal Theories and How To Unreify Them. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI'91)*, pages 1177–1182, Sydney, Australien, August 1991.
- [GCD07] Sten Govaerts, Nik Corthaut, and Eirik Duval. Mood-Ex-Machina: Towards Automation of Moody Tunes. In *Proceedings of the 8th International Conference on Music Information Retrieval*, Wien, Österreich, September 2007.
- [Gol05] E.B. Goldstein. *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*. Wadsworth Publishing, 2005.
- [Gon05] Roberto García González. *A Semantic Web approach to Digital Rights Management*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spanien, November 2005.
- [GR07] Frédéric Giasson and Yves Raimond. Music Ontology Specification [online]. Februar 2007. Available from: <http://musicontology.com/> [cited 12.03.2009].
- [Gra01] Gracenote. *Gracenote CDDB® SDK for Windows Technical Summary*. Gracenote, März 2001.
- [Gra09a] Gracenote. Gracenote:Frequently Asked Questions [online]. 2009. Available from: http://www.gracenote.com/company_info/FAQ/FAQs/ [cited 04.09.2009].

- [Gra09b] Gracenote. MusicID [online]. 2009. Available from: http://www.gracenote.com/business_solutions/music_id/ [cited 12.02.2009].
- [GsC05] Roberto García and Óscar Celma. Semantic Integration and Retrieval of Multimedia Metadata. In *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation*, Galway, Irland, November 2005.
- [Hai00] Kurt Haider. *Einführung in die Musiktheorie*. Peter Lang, 2000.
- [Hau09] Imke Hauswirth. *Lautheits- und Schärfe-Berechnung mit ArtemiS*, März 2009. 23.05.2009.
- [Hem01] Christoph Hempel. *Neue Allgemeine Musiklehre*. Serie Musik Atlantis - Schott, 2001.
- [Her08] Oscar Celma Herrada. *Music Recommendation and Discovery in the Long Tail*. PhD thesis, Department of Technology, Universitat Pompeu Fabra, Barcelona, Spanien, 2008.
- [HL06a] Marko Helen and Tommi Lahti. Query by Example AMethods for Audio Signals. In *Proceedings of the 7th Nordic*, pages 302–305, Rejkjavik, Island, Juni 2006.
- [HL06b] Jane Hunter and Suzanne Little. MPEG-7 Ontology [online]. November 2006. Available from: <http://metadata.net/mpeg7/> [cited 09.02.2009].
- [Hol05] Peter Holtz. *Was ist Musik? Subjektive Theorien Musikschoffender Künstler*. PhD thesis, Philosophischen Fakultät I der Friedrich-Alexander-Universität Erlangen-Nürnberg, 2005.
- [Hun01] Jane Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *Proceedings of the 1st International Semantic Web Working Symposium*, Stanford, USA, August 2001.
- [HW30] Alois Höfler and Alois Wenzl. *Psychologie*. Adamant Media Corporation, 2 edition, 1930.
- [Hym08] Spencer Hyman. The 20 minute tour... In *Digital radio summit 2008*, Tokyo, Japan, November 2008.
- [ian09] ian. BBC Backstage SPARQL Endpoint [online]. Juni 2009. Available from: <http://welcomebackstage.com/2009/06/bbc-backstage-sparql-endpoint/> [cited 22.08.2009].
- [Inf09] Information Management Systems Research Group. Music information retrieval [online]. 2009. Available from: <http://ims.dei.unipd.it/websites/cms/research/music-information-retrieval.html> [cited 29.08.2009].

- [IS05] Julius O. Smith III and Xavier Serra. PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation. Technical report, Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, 2005.
- [Jeh05] Tristan Jehan. *Creating Music by Listening*. PhD thesis, School of Architecture and Planning, Massachusetts Institute of Technology, September 2005.
- [JLY01] Jyh-Shing Roger Jang, Hong-Ru Lee, and Chia-Hui Yeh. Query by Tapping: A New Paradigm for Content-Based Music Retrieval from Acoustic Input. In *Advances in Multimedia Information Processing - PCM 2001*. Springer Berlin / Heidelberg, 2001.
- [Jon09] Richard Jones. Last.fm Radio Announcement [online]. März 2009. Available from: <http://blog.last.fm/2009/03/24/lastfm-radio-announcement> [cited 06.09.2009].
- [JR09] Kurt Jacobson and Yves Raimond. The Similarity Ontology [online]. Juli 2009. Available from: <http://grasstunes.net/ontology/similarity/0.2/musim.html#Similarity> [cited 28.08.2009].
- [JTh07] JThink Ltd. Jaudiotagger [online]. 2007. Available from: <http://www.jthink.net/jaudiotagger/index.jsp> [cited 19.08.2009].
- [KC08] Ian Knopke and Chris Cannam. *Sonic Annotator*, Juli 2008.
- [Kla99] Anssi Klapuri. Sound Onset Detection by Applying Psychoacoustic Knowledge. In *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing*, volume 6, pages 115–118, Washington, DC, USA, 1999.
- [KPS⁺08] Peter Knees, Tim Pohle, Markus Schedl, Dominik Schnitzer, and Klaus Seyerlehner. A Document-Centered Approach to a Natural Language Music Search Engine. In *ECIR*, pages 627–631, Glasgow, Scotland, GB, April 2008.
- [KSB⁺08] Hak Lae Kim, Simon Scerri, John G. Breslin, Stefan Decker, and Hong Gee Kim. The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies. In *Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*, pages 128–137, Berlin, Deutschland, 2008. Dublin Core Metadata Initiative.
- [Las09] Lastfm Ltd. Last.fm - What is fingerprinting? [online]. 2009. Available from: <http://www.last.fm/help/faq?category=The+Last.fm+Scrobbler#183> [cited 06.09.2009].
- [LD04] Jin Ha Lee and J. Stephen Downie. Survey of Music Information Needs, Uses, and Seeking Behaviours: Preliminary Findings. In *Proceedings of the 5th International Conference on Music Information Retrieval*, Barcelona, Spanien, Oktober 2004.

- [LDJ07] Jin Ha Lee, J. Stephen Downie, and M. Cameron Jones. Preliminary Analyses of Information Features Provided by Users for Identifying Music. In *Proceedings of the 8th International Conference on Music Information Retrieval*, Wien Österreich, September 2007.
- [Les06] Micheline Lesaffre. *Music Information Retrieval - Conceptual Framework, Annotation and User Behaviour*. PhD thesis, Faculty of Arts and Philosophy, Department of Art, Music and Theatre Sciences Ghent University, 2006.
- [Lid06] Thomas Lidy. Evaluation of New Audio Features and Their Utilization in Novel Music Retrieval Applications. Master's thesis, Institut für Softwaretechnik und Interaktive Systeme, Technischen Universität Wien, Wien, Österreich, Dezember 2006.
- [Lil08] Anita Shen Lillie. MusicBox: Navigating the space of your music. Master's thesis, School of Architecture and Planning, Massachusetts Institute of Technology, September 2008.
- [LML04] Micheline Lesaffre, Dirk Moelants, and Marc Leman. *Music Query: Methods, Models, and User Studies*, volume 13 of *Computing in Musicology*, chapter Spontaneous User Behavior in 'Vocal' Queries for Music-Information Retrieval, pages 129–146. CCARH and The MIT Press, 2004.
- [Log00] Beth Logan. Mel Frequency Cepstral Coefficients for Music Modelling. In *Proceedings of the 1st International Conference on Music Information Retrieval*, Plymouth, Massachusetts, USA, Oktober 2000.
- [LS01] Beth Logan and Ariel Salomon. A Music Similarity Function Based on Signal Analysis. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, Tokyo, Japan, 2001.
- [LVH05] Tuck W Leong, Frank Vetere, and Steve Howard. The serendipity shuffle. In *Proceedings of the 17th Australia conference on Computer-Human Interaction*, pages 1–4, Narrabundah, Australien, 2005. Computer-Human Interaction Special Interest Group (CHISIG) of Australia.
- [Mac09a] Macrovision. About Us [online]. 2009. Available from: http://www.allmusic.com/cg/amg.dll?p=amg&sql=32:amg/info_pages/a_about.html [cited 12.02.2009].
- [Mac09b] Macrovision. All Music Guide [online]. 2009. Available from: http://www.macrovision.com/products/online_stores_portals/data_licensing/amg_data_solutions.htm?tabContent=/products/online_stores_portals/data_licensing/amg_data_solutions/edition_8772.htm [cited 13.03.2009].
- [Mac09c] Macrovision. Coverage Statistics [online]. März 2009. Available from: http://www.macrovision.com/products/online_stores_portals/data_licensing/amg_data_solutions.

- htm?tabContent=/products/online_stores_portals/data_licensing/amg_data_solutions/edition_8780.htm [cited 13.03.2009].
- [Mac09d] Macrovision. LASSO [online]. 2009. Available from: http://www.macrovision.com/products/ce_manufacturers/metadata_media_recognition/lasso.htm [cited 12.02.2009].
- [MAG06] MAGIX. Frequently Asked Questions [online]. 2006. Available from: <http://www.freedb.org/en/faq.3.html> [cited 12.02.2009].
- [Mar04] MPEG-7 Overview [online]. Oktober 2004. Available from: <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm> [cited 04.02.2009].
- [Mar08a] Matija Marolt. A Mid-Level Representation for Melody-based Retrieval in Audio Collections. *Multimedia, IEEE Transactions on*, 10(8):1617–1625, Dezember 2008.
- [Mar08b] Luis Gustavo Pereira Marques Martins. *A Computational Framework for Sound Segregation in Music Signals*. PhD thesis, Faculdade de Engenharia da Universidade do Porto, September 2008.
- [May07a] Mayhem. FreeDB - MusicBrainz gateway now available! [online]. Juni 2007. Available from: <http://blog.musicbrainz.org/?p=273> [cited 12.02.2009].
- [May07b] Mayhem. The old RDF based web service has been deprecated [online]. Januar 2007. Available from: <http://blog.musicbrainz.org/?p=233> [cited 11.02.2009].
- [MCR08] Michela Magas, Michael Casey, and Christophe Rhodes. mHashup: Fast Visual Music Discovery via Locality Sensitive Hashing. In *Proceedings of the International Conference on Computer Graphics and Interactive Techniques 2008*, 2008.
- [Mei05] Klaus Meißner. Audio-Technik, -Formate und Schnittsysteme. In *Medien und Medienströme*, pages 1–82. TU Dresden, 2005.
- [MF06] Cory McKay and Ichiro Fujinaga. Musical Genre Classification: Is It Worth Pursuing and How Can It be Improved? In *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Kanada, Oktober 2006.
- [mis07] miscellaneous. Gleaning Resource Descriptions from Dialects of Languages (GRDDL) [online]. September 2007. Available from: <http://www.w3.org/TR/grddl/> [cited 21.08.2009].
- [mis08a] miscellaneous. APEv1 - Hydrogenaudio Knowledgebase [online]. Februar 2008. Available from: <http://wiki.hydrogenaudio.org/index.php?title=APEv1> [cited 03.02.2009].

- [mis08b] miscellaneous. APEv2 - Hydrogenaudio Knowledgebase [online]. Februar 2008. Available from: <http://wiki.hydrogenaudio.org/index.php?title=APEv2> [cited 03.02.2009].
- [mis08c] miscellaneous. The MusicBrainz XML Metadata Format [online]. 2008. Available from: <http://musicbrainz.org/doc/MusicBrainzXMLMetaData> [cited 12.02.2009].
- [mis08d] miscellaneous. XMLWebService [online]. Dezember 2008. Available from: <http://wiki.musicbrainz.org/XMLWebService> [cited 11.02.2009].
- [mis09a] miscellaneous. audioDB [online]. 2009. Available from: <http://omras2.org/audioDB> [cited 28.08.2009].
- [mis09b] miscellaneous. Feature vector database management system [online]. August 2009. Available from: <http://rvw.doc.gold.ac.uk/sullivan/cgi-bin/gitweb.cgi?p=audioDB.git;a=summary> [cited 28.08.2009].
- [mis09c] Grove Music Online [online]. 2009. Available from: <http://www.oxfordmusiconline.com/> [cited 26.05.2009].
- [mis09d] miscellaneous. How PUIDs Work [online]. Juni 2009. Available from: http://wiki.musicbrainz.org/How_PUIDs_Work [cited 18.08.2009].
- [mis09e] miscellaneous. Linking Open Data [online]. August 2009. Available from: <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/> [cited 20.08.2009].
- [mis09f] miscellaneous. List of Music Genome Project attributes [online]. August 2009. Available from: http://en.wikipedia.org/wiki/List_of_Music_Genome_Project_attributes [cited 06.09.2009].
- [mis09g] miscellaneous. List of Music Genome Project attributes by type [online]. August 2009. Available from: http://en.wikipedia.org/wiki/List_of_Music_Genome_Project_attributes_by_type [cited 06.09.2009].
- [mis09h] miscellaneous. Range (music) [online]. 2009. Available from: http://en.wikipedia.org/wiki/Pitch_range [cited 01.09.2009].
- [MKC06] Meinard Müller, Frank Kurth, and Michael Clausen. Aktuelle Aspekte des Music Information Retrieval. *Datenbank-Spektrum*, 18/2006:24–32, 2006.
- [MMWB05] Alistair Miles, Brian Matthews, Michael Wilson, and Dan Brickley. SKOS core: Simple Knowledge Organisation for the Web. In *Proceedings of the 2005 international conference on Dublin Core and metadata applications*, pages 1–9, Madrid, Spanien, 2005. Dublin Core Metadata Initiative.

- [MSS03] B. S. Manjunath, Phillippe Salembier, and Thomas Sikora. *Introduction to MPEG-7*. Wiley, 2003.
- [muf08] mufin GmbH. See music different with mufin vision [online]. 2008. Available from: <http://business.mufin.com/en/products/mufin-vision-innovative-visualization-of-digital-music-libraries/> [cited 03.05.2009].
- [muf09] mufin GmbH. mufin audioid - Musikererkennung und Musikidentifizierung [online]. 2009. Available from: <http://business.mufin.com/de/produkte/mufin-audioid-musikererkennung-und-musikidentifizierung/> [cited 05.08.2009].
- [Mus09] Muscle. MUSCLE Network of Excellence - Feature Extraction Tools for Audio [online]. 2009. Available from: http://www.ifs.tuwien.ac.at/mir/muscle/del/audio_tools.html [cited 05.08.2009].
- [Nie01] Helmut M. Niegemann. Sensorisches Gedächtnis [online]. 2001. Available from: http://www.stud.tu-ilmenau.de/~waskow/sensorisches_gedaechtnis.htm [cited 23.05.2009].
- [O’N06a] Dan O’Neill. ID3v1 - ID3.org [online]. Oktober 2006. Available from: <http://www.id3.org/ID3v1> [cited 03.02.2009].
- [O’N06b] Dan O’Neill. ID3v2Easy - ID3.org [online]. Dezember 2006. Available from: <http://www.id3.org/ID3v2Easy> [cited 23.02.2009].
- [O’N09] Dan O’Neill. Home - ID3.org [online]. Januar 2009. Available from: <http://www.id3.org/Home> [cited 02.02.2009].
- [Ope09a] OpenLink Software. RDF Data Access and Data Management [online]. 2009. Available from: <http://docs.openlinksw.com/virtuoso/rdfandsparql.html> [cited 28.08.2009].
- [Ope09b] OpenLink Software. Virtuoso Open-Source Edition [online]. 2009. Available from: <http://virtuoso.openlinksw.com/dataspace/dav/wiki/Main/> [cited 28.08.2009].
- [O’S87] Douglas O’Shaughnessy. *Speech communication: human and machine*. Addison-Wesley, 1987.
- [Pac05] Francois Pachet. Knowledge Management and Musical Metadata. In *Encyclopedia of Knowledge Management*, pages 672–677. Idea Group, September 2005.
- [Pam01] Elias Pampalk. Island of Music - Analysis, Organisation and Visualisation of Music Archives. Master’s thesis, Institut für Softwaretechnik und Interaktive Systeme, Technische Universität Wien, 2001.

- [Pam06] Elias Pampalk. *Computational Models of Music Similarity and their Application in Music Information Retrieval*. PhD thesis, Fakultät für Informatik, Technischen Universität Wien, 2006.
- [Par75] Denys Parsons. *The Directory of Tunes and Musical Themes*. S. Brown, 1975.
- [Par07] Aliaksandr V. Paradzinets. *Variable Resolution Transform-based Music Feature Extraction and their Applications for Music Information Retrieval*. PhD thesis, Ecole Centrale de Lyon, 2007.
- [PC00] Francois Pachet and Daniel Cazaly. A Taxonomy of Musical Genres. In *In Proceedings of the 1st Conference of Content-Based Multimedia Information Access (RIAO)*, Paris, France, April 2000.
- [PC09] Aliaksandr Paradzinets and Liming Chen. Use of Variable Resolution Transform for Musical Descriptor Rxttraction. Technical Report RR-LIRIS-2009-029, LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon, Juni 2009. Available from: <http://liris.cnrs.fr/publis/?id=4225>.
- [Pee04] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, Icrum, 2004.
- [Pfl08] Bernhard Pflugfelder. Discriminant Analysis of Three Rhythmic Descriptors in Musical Genre Classification. Master's thesis, Institut für Softwaretechnik und Interaktive Systeme, Technische Universität Wien, 2008.
- [PHC06] Aliaksandr Paradzinets, Hadi Harb, and Liming Chen. Use of continuous wavelet-like transform in automated music transcription. In *Proceedings of 14th European Signal Processing Conference*, Florenz, Italien, September 2006.
- [PR08] Alexandra Passant and Yves Raimond. Combining Social Music and Semantic Web for Music-Related Recommender Systems. In *Social Data on the Web Workshop of 7th International Semantic Web Conference*, Karlsruhe, Deutschland, Oktober 2008.
- [Qpa09] Qpaly. Rhythmus (Musik) [online]. April 2009. Available from: http://de.wikipedia.org/wiki/Rhythmus_%28Musik%29 [cited 01.09.2009].
- [RA07a] Yves Raimond and Samer Abdallah. The Event Ontology [online]. Oktober 2007. Available from: <http://motools.sourceforge.net/event/event.html> [cited 03.02.2009].
- [RA07b] Yves Raimond and Samer Abdallah. The Timeline Ontology [online]. Oktober 2007. Available from: <http://motools.sourceforge.net/timeline/timeline.html> [cited 03.02.2009].

- [Rai08a] Yves Raimond. *A Distributed Music Information System*. PhD thesis, Queen Mary, University of London, November 2008.
- [Rai08b] Yves Raimond. Echonest Analyze XML to Music Ontology RDF [online]. Juli 2008. Available from: <http://dbtune.org/echonest/> [cited 21.08.2009].
- [RASG07] Yves Raimond, Samer Abdallah, Mark Sandler, and Frederick Giasson. The Music Ontology. In *Proceedings of the 8th International Conference on Music Information Retrieval*, pages 417–422, Wien, Österreich, September 2007.
- [Roe00] Juan G. Roederer. *Physikalische und psychoakustische Grundlagen der Musik*. Springer, 2000.
- [Ros73] Eleanor Rosch. Natural categories. *Cognitive Psychology*, 4:328–350, 1973.
- [RP08] Patrick Rabbat and François Pachet. Direct and Inverse Inference in Music Databases: How to Make a Song Funk? In *Proceedings of the Ninth International Conference on Music Information Retrieval*, Philadelphia, USA, September 2008.
- [Sau07] Franz Sauter. Musiktheorie online [online]. 2007. Available from: <http://www.tonalemusik.de/> [cited 27.05.2009].
- [Sau09] Leo Sauermann. Extractors – aperture [online]. Juli 2009. Available from: <http://sourceforge.net/apps/trac/aperture/wiki/Extractors> [cited 19.08.2009].
- [SCBG08] Mohamed Sordo, Oscar Celma, Martin Blech, and Enric Guaus. The Quest for Musical Genres: Do the Experts and the Wisdom of Crowds Agree? In *Proceedings of the 9th International Conference on Music Information Retrieval*, Philadelphia, USA, September 2008.
- [Sch08] Erick Schonfeld. Pandora Hits 20 Million Registered Users [online]. Dezember 2008. Available from: <http://www.techcrunch.com/2008/12/19/pandora-hits-20-million-registered-users-via-twitter/> [cited 06.09.2009].
- [Sep01] Jarno Seppänen. Tatum Grid Analysis of Musical Signals. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, New York, USA, Oktober 2001.
- [Ser08] Xavier Serra. Sound and Music Description. Presentation at Music Technology Group, Universitat Pompeu Fabra, Barcelona, 2008.
- [SF00] Eleanor Selfridge-Field. What Motivates a Musical Query? In *Proceedings of the 1st International Conference on Music Information Retrieval*, Plymouth, Massachusetts, USA, Oktober 2000.

- [SKW08] Malcolm Slaney, Kilian Weinberger, and William White. Learning a Metric for Music Similarity. In *Proceedings of the 9th International Conference on Music Information Retrieval*, Philadelphia, USA, September 2008.
- [SLS08] Rebecca Stewart, Mark Levy, and Mark Sandler. 3D Interactive Environment for Music Collection Navigation. In *Proceedings of the 11th International Conference on Digital Audio Effects*, Espoo, Finland, September 2008.
- [SW08] Markus Schedl and Gerhard Widmer. Automatically Detecting Members and Instrumentation of Music Bands Via Web Content Mining. *Adaptive Multimedial Retrieval: Retrieval, User, and Semantics: 5th International Workshop, AMR 2007, Paris, France, July 5-6, 2007 Revised Selected Papers*, 4918/2008:122–133, 2008. doi:http://dx.doi.org/10.1007/978-3-540-79860-6_10.
- [TBL08] Douglas Turnbull, Luke Barrington, and Gert Lanckriet. Five Approaches to Collecting Tags for Music. In *Proceedings of the 9th International Conference on Music Information Retrieval*, Philadelphia, USA, September 2008.
- [TBTL07] Douglas Turnbull, Luke Barrington, David Torres, and Gert Lanckriet. Towards Musical Query-by-Semantic-Description using the CAL500 Data Set. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, Amsterdam, Niederlande, 2007.
- [TC07] Chrisa Tsinaraki and Panagiotis Polydoros and Stavros Christodoulakis. Interoperability Support between MPEG-7/21 and OWL in DS-MIRF. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 19(2):219–232, 2007.
- [Ter98] Ernst Terhardt. *Akustische Kommunikation*. Springer, Berlin, 1998.
- [Ter08] Susan Terge. Prototype on Speech and Music Retrieval Systems with Vocal Query Interface. Technical report, Queen Mary University of London, Januar 2008.
- [The08] The Echo Nest Corporation. analyze [online]. 2008. Available from: <http://the.echonest.com/analyze/> [cited 05.08.2009].
- [The09] The Echo Nest Corporation. The Echo Nest : API : Overview [online]. 2009. Available from: <http://developer.echonest.com/pages/overview> [cited 19.08.2009].
- [Tim05] Christian Timmerer. Information Technology–MPEG-21 Multimedia Framework [online]. 2005. Available from: <http://mpeg-21.itec.uni-klu.ac.at/cocoon/mpeg21/> [cited 09.02.2009].
- [Tim07] Christian Timmerer. The MPEG-21 Multimedia Framework [online]. Juni 2007. Available from: <http://multimediacommunication>.

- blogspot.com/2007/06/mpeg-21-multimedia-framework.html [cited 09.02.2009].
- [TLBL07] Douglas Turnbull, Ruoran Liu, Luke Barrington, and Gert Lanckriet. A Game-Based Approach for Collecting Semantic Annotations of Music. In *Proceedings of the 8th International Conference on Music Information Retrieval*, Wien, Österreich, September 2007.
- [TPC04] Chrisa Tsinaraki, Panagiotis Polydoros, and Stavros Christodoulakis. Interoperability support for Ontology-based Video Retrieval Applications. In *Proceedings of 3rd International Conference on Image and Video Retrieval*, volume 3115/2004, pages 582 – 591, Juli 2004.
- [TS92] Jean Tague-Sutcliffe. The pragmatics of information retrieval experimentation, revisited. *Inf. Process. Manage.*, 28(4):467–490, 1992. doi:[http://dx.doi.org/10.1016/0306-4573\(92\)90005-K](http://dx.doi.org/10.1016/0306-4573(92)90005-K).
- [Tza02] George Tzanetakis. *Manipulation, Analysis and Retrieval Systems for Audio Signals*. PhD thesis, Computer Science Department, Princeton University, 2002.
- [Vas01] Pantelis N. Vassilakis. *Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance*. PhD thesis, University of California, 2001.
- [Vas05] Pantelis N. Vassilakis. Auditory Roughness as a Means of Musical Expression. *Selected Reports in Ethnomusicology*, 7 (Perspectives in Musicology):119–144, 2005.
- [WF05] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.
- [WFC07] Ge Wang, Rebecca Fiebrink, and Perry R. Cook. Combining Analysis and Synthesis in the ChuK Programming Language. In *Proceedings of the 2007 International Computer Music Conference (ICMC)*, Kopenhagen, Dänemark, 2007.
- [whe08] wheeler. TagLib Audio Meta-Data Library [online]. 2008. Available from: <http://developer.kde.org/~wheeler/taglib.html> [cited 19.08.2009].
- [Whi05] Brian A. Whitman. *Learning the Meaning of Music*. PhD thesis, School of Architecture and Planning, Massachusetts Institute of Technology, 2005.
- [Wie17] Theodor Wiehmayer. *Musikalische Rhythmik und Metrik*. Heinrichshofen's Verlag, 1917.
- [Wim04] Elisabeth Wimmer. MUSIK–Inspiration oder Berieselung. Master's thesis, Private Pädagogische Hochschule der Diözese Linz, 2004.

- [WZO07] Bin Wei, Chengliang Zhang, and Mitsunori Ogihara. Keyword Generation for Lyrics. In *Proceedings of the 8th International Conference on Music Information Retrieval*, Wien, Österreich, September 2007.
- [Xip05] Xiph.org. Ogg Vorbis Documentation [online]. 2005. Available from: <http://www.xiph.org/vorbis/doc/v-comment.html> [cited 03.02.2009].
- [Xip08] Xiph.org. M3F - XiphWiki [online]. Juni 2008. Available from: <http://wiki.xiph.org/index.php/M3F> [cited 03.02.2009].
- [Xip09] Xiph.org. VorbisComment - XiphWiki [online]. Januar 2009. Available from: <http://wiki.xiph.org/index.php/VorbisComment> [cited 03.02.2009].
- [Zhu08] Xiaojin Zhu. Semi-Supervised Learning Literature Survey. online; http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey_7_19_2008.pdf, July 2008. Survey.